2017

Assessing non-inferiority via risk difference in one-to-many propensity-score matched studies

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

ASSESSING NON-INFERIORITY VIA RISK DIFFERENCE IN ONE-TO-MANY PROPENSITY-SCORE MATCHED STUDIES

by

JEREMIAH PEREZ

B.S.B.A, University of Florida, 2008
B.A., University of Florida, 2008
M.S., University of Central Florida, 2012

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2017

Approved by

First Reader ______________________________

Joseph M. Massaro, Ph.D.
Professor of Biostatistics

Second Reader ______________________________

Ralph D'Agostino, Ph.D.
Professor of Mathematics and Statistics

Third Reader ______________________________

Michael P. LaValley, Ph.D.
Professor of Biostatistics

DEDICATION

I would like to dedicate this work to my parents for their endless love, support, and encouragement.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my primary advisor Professor Massaro for the continuous support of my Ph.D. study and research. I would also like to thank the rest of my dissertation committee: Professor D'Agostino, Professor LaValley, Professor Cabral, and Professor Milton for their insightful comments, questions, advice, and encouragement.

ASSESSING NON-INFERIORITY VIA RISK DIFFERENCE IN ONE-TO-MANY PROPENSITY-SCORE MATCHED STUDIES

JEREMIAH PEREZ

Boston University Graduate School of Arts and Sciences, 2017

Major Professor: Joseph M. Massaro, Professor of Biostatistics

ABSTRACT

Non-inferiority tests are well developed for randomized parallel group trials where the control and experimental groups are independent. However, these tests may not be appropriate for assessing non-inferiority in correlated one-to-many matched data. We propose a new statistical test that extends Farrington-Manning's (FM) test to the case where many ($\geq$1) control subjects are matched to each experimental subject. We conducted a Monte Carlo simulation study to compare the size and power of the proposed test with tests developed for clustered one-to-one matched pair data and tests based on generalized estimating equations (GEE). For various correlation patterns, the sizes of tests developed for clustered matched pair data and GEE-based tests are inflated when applied to the case where many control subjects are matched to each experimental subject. The size of the proposed test, on the other hand, is close to the nominal level for a variety of correlation patterns.

There is a debate in the literature regarding whether or not statistical tests appropriate for independent samples can be used to assess the statistical significance of treatment effects in propensity-score matched studies. We used Monte Carlo simulations to examine the effect on assessing non-inferiority via risk difference when a method for

independent samples (i.e. FM test) is used versus when a method for correlated matched samples is used in propensity-score one-to-many matched studies. If propensity-score matched samples are well-matched on baseline covariates and contain almost all of the experimental treated subjects, a method for correlated matched samples is preferable with respect to power and Type I error than a method for independent samples.

Sometimes there are more experimental subjects to choose from for matching than control subjects. We conducted a Monte Carlo simulation study to compare the size and power of the previously mentioned tests when many (≥ 1) experimental subjects are matched to each control subject. In this case, the Nam-Kwon test for clustered data performs the best in controlling the type I error rate for a variety of correlation patterns. Therefore, the appropriate non-inferiority test to use for correlated matched data depends, in part, on the sample size allocation of subjects.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ILLUSTRATIONS

1. NON-INFERIORITY TESTS FOR ONE-TO-MANY MATCHED DATA

1.1. Introduction

Non-inferiority clinical trials are often used to show that a new experimental treatment is not materially worse than an active control by a predetermined amount (i.e. non-inferiority margin) with respect to effectiveness. This leads to the following hypothesis test for non-inferiority:

$$\mathrm{H}_0\colon p_C - p_E \leq \delta_0 \quad \text{vs.} \quad \mathrm{H}_1\colon p_C - p_E > \delta_0$$

where p_C and p_E are the true unknown event probabilities for the control and experimental procedures, respectively, and $\delta_0 < 0$ is the non-inferiority margin. Without loss of generality, we assume that a smaller success probability denotes greater efficacy. There are several well developed statistical methods for making inferences about absolute risk differences for randomized parallel-group non-inferiority trials where the two treatment groups are independent. For example, the Farrington-Manning (1990) test, which is based on the restricted maximum likelihood estimation, is frequently used in randomized trials to assess non-inferiority.

In non-randomized two-treatment trials, patients are assigned to the treatment groups of interest in a non-randomized manner, which increases the likelihood of imbalance on baseline characteristics between the treatment groups. This imbalance on baseline characteristics may lead to biased estimates of treatment effect. Matching treated patients to control patients on baseline covariate information (through a propensity score for example) is a common technique to reduce the bias in treatment effect that is caused by the lack of randomization. After matching, outcomes can be compared between treatment groups on the matched data. However, matched treatment groups may not be independent. Thus, the analysis of matched data requires statistical methods that account for the possible correlation between matched subjects.

There are several statistical tests available to assess non-inferiority via risk difference for 1:1 matched (i.e. matched-pair) data. Nam (1997) developed a test statistic to establish non-inferiority for 1:1 matched data based on the restricted maximum likelihood estimation and assuming matched-pairs are independent. However, this test may not be appropriate for clustered matched-pair data where pairs within a cluster are correlated (e.g., each experimental in a given non-randomized clinical trial might be matched to active control patients only from the same study center; in this case, study center would be considered the cluster). Therefore, Nam and Kwon (2009) proposed an adjusted score test that applies the intra-class correlation coefficient (*ICC*) to account for the clustering. Durkalski et al (2003) also proposed a non-inferiority test for clustered 1:1 matched data based on the method of moments (*MOM*). Nam and Kwon (2009) also proposed a test to evaluate non-inferiority for clustered 1:1 matched data based on a test statistic developed by Obuchowski (*O*) (1998). Nam and Kwon (2009) compared these three methods (i.e. *ICC, MOM, O*) for clustered matched-pair data and concluded that all of the tests are appropriate for practice; however, they came to this conclusion despite the increase in empirical Type I error in several situations that they investigated. The simulation study of Nam and Kwon (2009) found that when the number of clusters is small or when the event rate is low, the test based on method of moments and Obuchowski's test are anti-conservative (i.e. empirical Type I error rates are inflated).

Also, the empirical Type I error rate for the ICC-adjusted statistic tend to be inflated when there is large variability of cluster sizes. Yang et al (2012) also performed an extensive Monte Carlo simulation study to examine the performance of these tests in assessing non-inferiority for clustered 1:1 matched data under various scenarios (e.g. different number of clusters, cluster sizes, event rates, and correlations). Yang et al (2012) found that one should consider all of these different scenarios and issues when choosing an appropriate test statistic. They found that the ICC-adjusted test statistic is generally recommended to effectively control the nominal type I error rate when there is constant or small variability of cluster sizes. For a greater number of clusters, the other test statistics (i.e. method of moments and Obuchowski statistic) maintain the nominal level reasonably well and have higher power. It is unclear how well these non-inferiority tests perform in the setting of one-to-many matching.

A common feature for all aforementioned test statistics is that no assumptions are made about the correlation structure. Note that one could potentially use generalized estimating equations (GEE) for comparing correlated proportions by pre-specifying a working correlation structure. However, in practice, little is known about the appropriate correlation structure. Therefore, there are many settings where a simple non-inferiority test that have minimal assumptions on the structure of the correlation is preferred when analyzing matched data. GEE methods have been used in the analysis of correlated longitudinal (i.e. repeated-measures) binary data in non-inferiority trials (Lee, 2008). Under the context of longitudinal data, simulation studies have demonstrated that the significance level of non-inferiority tests based on GEE is generally less than 10% above a nominal level of 0.05 for various settings (Lee, 2008). GEE methods has been used to detect differences between treatments for matched-pair binary data (SAS, 2012), but little work has been done in applying GEE methods to specifically assess non-inferiority via risk difference for one-to-many matched-data.

When there are large numbers of control individuals, it is sometimes possible to get multiple good matches for each treated individual. However, it is unclear how to assess non-inferiority via risk difference when multiple control subjects are matched to each treated subject (Austin, Statistical Criteria for Selecting the Optimal Number of Untreated Subjects Matched to Each Treated Subject When Using Many-to-One Matching on the Propensity Score, 2010). Selecting multiple controls for each treated individual will generally increase bias of the estimated treatment effect, but on the other hand, utilizing multiple matches can increase precision and power than does simple 1:1 matching (Stuart, 2010). One way to reduce bias in one-to-many matched studies is by not fixing the number of controls matched to each treated subject but rather allowing it to vary from one matched set to another (Ming & Rosenbaum, 2000). Thus, there is a need to specifically develop non-inferiority statistical tests for one-to-many matched studies that may use fixed-ratio or variable-ratio matching.

Section 2 presents a general discussion of how existing non-inferiority tests for clustered one-to-one matched data can be applied to one-to-many matched study designs. Section 3 presents a new non-inferiority test statistic for analyzing one-to-many matched data that uses the idea of the variance inflation factor (VIF) and Obuchowski's covariance estimation to adjust the Farrington-Manning test statistic in order to keep

Type I error close to the nominal level. Section 4 compares the new statistical test with existing methods in terms of empirical Type I error rates and power via Monte Carlo simulation. Section 5 provides a discussion.

1.2. Application of existing non-inferiority tests to analyze one-to-many matched data

We are interested in establishing non-inferiority of a new experimental treatment versus an active control in a non-randomized setting using matched data. A *matched set* is created when one subject who received the experimental procedure is matched on baseline covariate information to one or more subjects who received the control procedure. Let *K* be the total number of matched sets. Let n_k be the number of subjects from the control group that are matched to subject *k* from the experimental group, where *k*=1, …, *K*. N_C and N_E are the total sample sizes for the control and experimental treatment groups after matching, respectively. Therefore, $N_E = K$ and $N_C = \sum_{k=1}^{K} n_k$. Note that for all *k*, n_k will be constant under fixed-ratio matching, and will vary under variable-ratio matching. Let x_{Ck} and x_{Ek} be the number of events in the control group and experimental group, respectively, in the *k*th matched set. Note, since there is only one subject treated with the experimental procedure in each matched set, then x_{Ek} = 0 or 1.

Illustration 1 shows how we can easily present 1:n_k matched data as *clustered* 1:1 matched data. We propose treating each 1:n_k matched set as a *cluster* of 1:1 matched data by simply pairing the binary response of the experimental subject to the binary response of *each* matched control. For example, as seen in Illustration 1, a 1-to-3 matched set could be treated as a *cluster* of *three* 1-to-1 matched data.

Illustration 1. An example of the data structure of a 1:n_k matched study design treated as clustered 1:1 matched data.

Experimental Response	Control Response	
E_1	C_{11}	Cluster of *three* 1-to-1 matched data ($n_1 = 3$)
E_1	C_{12}	
E_1	C_{13}	
E_2	C_{21}	Cluster of *two* 1-to-1 matched data ($n_2 = 2$)
E_2	C_{21}	

Let E_k be the binary response (i.e. 1 or 0 as event or no event, respectively) of the kth experimental treated subject. Let C_{kj} be the binary response of jth control subject who was matched to the kth experimental treated subject, where j=1, 2…, n_k.

When our data is treated as matched-pair data, there are four possible pairs of responses, i.e. (1,1), (1,0), (0,1), and (0,0), where the first element is the response to the control procedure and the second element is the response to the experimental procedure. Let a_k, b_k, c_k, d_k be the observed frequencies of (1,1), (1,0), (0,1), and (0,0), respectively, in cluster *k*. Clearly, a_k and d_k are frequencies of concordant pairs, and b_k and c_k are frequencies of discordant pairs. The actual number of events in the experimental group is adjusted when we treat our 1:n_k data as *clustered* 1:1 matched data. Let $x_{Ek}^{Clustered} = x_{Ek}n_k$ be the "adjusted" number of events in the experimental group in the *k*th cluster. By treating each 1:n_k matched set as a *cluster* of 1:1 matched data, it may be possible to apply existing non-inferiority tests for *clustered* 1-to-1 matched data to 1-to-n_k matched study designs.

1.2.1 Test based on method of moments

Durkalski et al (2003) proposed a non-inferiority test statistic for clustered matched pair data based on method of moments,

$$Z_{MOM} = \sum_{k=1}^{K}\left(\frac{b_k-c_k}{n_k} - \delta_0\right) / \left[\sum_{k=1}^{K}\left(\frac{b_k-c_k}{n_k} - \delta_0\right)^2\right]^{\frac{1}{2}} \quad (1)$$

For a large number of clusters, the statistic (1) is approximately normal with mean zero and variance one under H_0. We reject H_0 against H_1 if $Z_{MOM} > z_{1-\alpha}$ where $z_{1-\alpha}$ is the upper 100(1-α) percentage point of the standard normal distribution.

1.2.2 ICC-adjusted test based on restricted MLE

If subjects within a cluster are positively correlated, test statistics ignoring the positive correlation are inflated and p-values are distorted downward, resulting in falsely rejecting non-inferiority. Eliasziw and Donner (1991) adjusted McNemar's test by utilizing a variance inflation factor (or design effect) to account for the underestimated variance resulting from clustering. Nam (1997) proposed a score statistic to test non-inferiority based on restricted maximum likelihood estimation (RMLE) for 1:1 matched binary data. Using a similar approach to adjust McNemar's test statistic for matched-pair data, Nam and Kwon (2009) adjusted the score statistic to test non-inferiority for clustered 1:1 matched-pair binary data,

$$Z_{ICC} = \left[\sum_{k=1}^{K}(b_k - c_k) - N_C\delta_0\right] / \left[N_C(\tilde{p}_{10} + \tilde{p}_{01} - \delta_0^2)(1 + (\widehat{m} - 1)\hat{\rho})\right]^{1/2} \quad (2)$$

where $\tilde{p}_{01}$ is the RMLE of p_{01}(the response probabilities for the pair (0,1)) and

$$\tilde{p}_{10} = \tilde{p}_{01} + \delta_0, \qquad \tilde{p}_{01} = \frac{\left\{-\tau + (\tau^2 - 4\upsilon\gamma)^{\frac{1}{2}}\right\}}{2\upsilon},$$

$\upsilon = 2\mathrm{N_c}$, $\tau = (2N_c + \sum_{k=1}^{K}(c_k - b_k))\delta_0 - \sum_{k=1}^{K}(b_k + c_k)$,

and $\gamma = -\delta_0(1-\delta_0)\sum_{k=1}^{K} c_k$

In statistic (2), $(1 + (\widehat{m} - 1)\hat{\rho})$ is the called the variance inflation factor (VIF), or the design effect (DE), where $\widehat{m}$ is the adjusted mean number of discordant pairs and $\hat{\rho}$ is a consistent estimator of the intra-class correlation coefficient (ICC). Nam and Kwon (2009) employed the analysis of variance (ANOVA) estimator of the intra-class correlation coefficient (ICC) to calculate VIF:

$S_k = b_k + c_k$, $K_d = \sum_{k=1}^{K} I_k$ where $I_k = \begin{cases} 0, & S_k = 0 \\ 1, & S_k \geq 1 \end{cases}$

$\bar{S} = \sum_{k=1}^{K} S_k / K_d$, $S_0 = \bar{S} - \{\sum_{k=1}^{K}(S_k - \bar{S})^2 - (K - K_d)\bar{S}^2\}/\{K_d(K_d - 1)\bar{S}\}$

$\widehat{m} = S_0 - K_d(S_0 - \bar{S})$

BMS = $\{\sum_{k=1}^{K} \frac{(b_k - S_k\bar{p})^2}{S_k}\}/K_d$ where $\bar{p} = \sum_{k=1}^{K} b_k \,/ \sum_{k=1}^{K} S_k$

WMS = $\sum_{k=1}^{K}((b_k\, c_k)/S_k)\,/\{K_d(\bar{S} - 1)\}$

$\hat{\rho}$ = (BMS-WMS) / {BMS + (S_0-1)WMS}

If $S_k = 0$, then the *k*th component of WMS is undefined and excluded from analysis. When all S_k's are zero across clusters, the $\hat{\rho}$ is not estimable. For a large number of clusters, the ICC-adjusted test statistic are distributed asymptotically normal mean zero and variance one.

1.2.3 Test based on Obuchowski's method

Based on sampling techniques, Obuchowski (1998) proposed a test statistic which takes into account correlation within a cluster. Nam and Kwon (2009) extended

Obuchowski's statistic to test for non-inferiority in clustered matched-pair data and proposed

$$\mathrm{Z_O} = (\hat{p}_C - \hat{p}_E - \delta_0)/\left[\widehat{\mathrm{Var}}(\hat{p}_C - \hat{p}_E - \delta_0)\right]^{\frac{1}{2}} \tag{3}$$

where $\widehat{\mathrm{Var}}(\hat{p}_C - \hat{p}_E - \delta_0) = \mathrm{Var}(\hat{p}_C) + \mathrm{Var}(\hat{p}_E) - 2\mathrm{Cov}(\hat{p}_C, \hat{p}_E)$. When writing the following variance formulas, we use the notation that was used earlier to describe how $1{:}n_k$ matched data can be treated as clustered matched-pair data:

$$\widehat{\mathrm{Var}}(\hat{p}_C) = \frac{K}{K-1}\left\{\frac{1}{\left(\sum_{k=1}^{K} n_k\right)^2}\sum_{k=1}^{K}(x_{Ck} - n_k\bar{p}_C)^2\right\},$$

$$\widehat{\mathrm{Var}}(\hat{p}_E) = \frac{K}{K-1}\left\{\frac{1}{\left(\sum_{k=1}^{K} n_k\right)^2}\sum_{k=1}^{K}\left(x_{Ek}^{Clustered} - n_k\bar{p}_E\right)^2\right\},$$

$$\widehat{\mathrm{Cov}}(\hat{p}_C, \hat{p}_E) = \frac{K}{K-1}\left\{\frac{1}{\left(\sum_{k=1}^{K} n_k\right)^2}\sum_{k=1}^{K}(x_{Ck} - n_k\bar{p}_C)\left(x_{Ek}^{Clustered} - n_k\bar{p}_E\right)\right\}$$

$$\hat{p}_C = \frac{1}{\sum_{k=1}^{K} n_k}\sum_{k=1}^{K} x_{Ck}, \qquad \hat{p}_E^{Clustered} = \frac{1}{\sum_{k=1}^{K} n_k}\sum_{k=1}^{K} x_{Ek}^{Clustered},$$

$$\bar{p}_C = \frac{\hat{p}_C + \hat{p}_E^{Clustered}}{2} + \frac{\delta_0}{2}, \qquad \bar{p}_E = \frac{\hat{p}_C + \hat{p}_E^{Clustered}}{2} - \frac{\delta_0}{2}$$

For a large number of clusters, statistic (3) is approximately a standard normal variate. We reject the null hypothesis at α when $\mathrm{Z_O} \geq Z_{1-\alpha}$.

1.2.4. Non-inferiority tests based on GEE

One could potentially use generalized estimating equations (GEE) for comparing correlated proportions when the data are sampled in clusters. However, it is unclear how well GEE methods perform when assessing non-inferiority based on risk difference for one-to-many matched data. In order to estimate the treatment effect, one could run a GEE model with the dichotomous outcome as the dependent variable and treatment group as the independent variable. To obtain an estimate of the risk difference, one could use an identity link function and assume a binomial distribution for the outcome in the GEE model. However, the identity link does not ensure that the model produces valid probability estimates. Errors may result when fitting such models depending on the model and the data (SAS, 2012). It is also common to assume a Poisson distribution for the outcome, if the binomial model for the risk difference fails to converge (Spiegelman & Hertzmark, 2005). To account for the possible correlation among matched subjects, one could specify an independent, exchangeable, or unstructured correlation structure in the GEE model. The coefficient of the treatment group variable in the GEE model will be the estimated risk difference β. Using the estimated risk difference and its standard error, we are able to derive a test statistic to assess non-inferiority, given a pre-specified working correlation structure s, where s={independence, exchangeable, unstructured} (Mascha & Sessler, 2011):

$$Z_s^{GEE} = (\beta - \delta_0)/SE(\beta) \tag{4}$$

1.3. Proposed extension of Farrington-Manning's test

In this section we propose a statistic that does not require us to treat one-to-many matched data as clustered matched-pair data. The proposed test statistic adjusts the Farrington-Manning (1990) test to account for the potential correlation within matched sets. Unlike GEE methods, this new method makes no assumptions about the correlation structure.

1.3.1 Farrington-Manning test for analyzing independent (non-matched) data

Farrington and Manning (1990) proposed the following non-inferiority test statistic assuming independent treatment groups based on the restricted maximum likelihood estimation

$$Z_{FM} = \frac{\hat{p}_C - \hat{p}_E - \delta_0}{\sqrt{\widehat{\mathrm{Var}}(\hat{p}_C - \hat{p}_E - \delta_0)_{FM}}} \tag{5}$$

where

$$\widehat{\mathrm{Var}}(\hat{p}_C - \hat{p}_E - \delta_0)_{FM} = \frac{\widetilde{p_C}(1 - \widetilde{p_C})}{N_C} + \frac{\widetilde{p_E}(1 - \widetilde{p_E})}{N_E}$$

$$\hat{p}_C = \frac{1}{N_c}\sum_{k=1}^{K} x_{Ck}\,, \qquad \hat{p}_E = \frac{1}{N_E}\sum_{k=1}^{K} x_{Ek},$$

$$\tilde{p}_C = 2u_{fm}\cos\left(w_{fm}\right) - b_{fm}/3a_{fm}, \qquad \tilde{p}_E = \tilde{p}_C - \delta_0$$

where

$$\theta = \frac{\mathrm{N_E}}{N_C}, \qquad a_{fm} = 1 + \theta, \qquad b_{fm} = -\left(1 + \theta + \hat{p}_C + \theta\hat{p}_E + \delta_0(\theta + 2)\right),$$

$$c_{fm} = (-\delta_0)^2 + \delta_0(2\hat{p}_C + \theta + 1) + \hat{p}_C + \theta\hat{p}_E, \quad \mathrm{d_{fm}} = -\delta_0 * (1 + \delta_0)\hat{p}_C,$$

$$u_{fm} = sign\left(v_{fm}\right)\sqrt{\left(\frac{b_{fm}^2}{\left(3a_{fm}\right)^2} - \frac{c_{fm}}{3a_{fm}}\right)},$$

$$v_{fm} = \left(\frac{\mathrm{b}_{\mathrm{fm}}^3}{(3\mathrm{a}_{\mathrm{fm}})^3}\right) - \left(\frac{\mathrm{b}_{\mathrm{fm}}\mathrm{c}_{\mathrm{fm}}}{6\mathrm{a}_{\mathrm{fm}}^2}\right) + \left(\frac{\mathrm{d}_{\mathrm{fm}}}{2\mathrm{a}_{\mathrm{fm}}}\right), \quad \mathrm{w}_{\mathrm{fm}} = (\pi + \cos^{-1}\left(\frac{v_{fm}}{u_{fm}^3}\right))/3.$$

$\tilde{p}_C$ and $\tilde{p}_E$ are the restricted MLEs of $\mathrm{p_C}$ and $\mathrm{p_E}$ under the null hypothesis that the risk difference equals δ_0. The test statistic $\mathrm{Z_{FM}}$ is asymptotically standard normal under the null hypothesis.

1.3.2 Adjusted Farrington-Manning test for analyzing one-to-many matched data

The Farrington-Manning test statistic is not applicable for matched data because it does not account for the potential correlation within subjects who are matched. However, test statistic (5) can be adjusted properly by

1) using a variance inflation factor (VIF) to account for the possible correlation among controls belonging to the same matched set (intra-class correlation), and
2) taking into account the correlation between matched experimental and control subjects (inter-class correlation) by applying Obuchowski’s method to estimate the covariance between the control and experimental groups’ event rates.

The following adjusted test statistic is appropriate for testing non-inferiority when each treated subject is matched to one or more control subjects:

$$Z_{AFM} = \frac{\hat{p}_C - \hat{p}_E - \delta_0}{\sqrt{\widehat{\mathrm{Var}}_{AFM}(\hat{p}_C - \hat{p}_E - \delta_0)}} \tag{6}$$

where

$$\widehat{\mathrm{Var}}_{\mathrm{AFM}}(\hat{p}_C - \hat{p}_E - \delta_0) = \frac{\tilde{p}_C(1-\tilde{p}_C)}{N_C} VIF_C + \frac{\tilde{p}_E(1-\tilde{p}_E)}{N_E} - 2Cov(\hat{p}_C, \hat{p}_E)$$

where $\tilde{p}_C$ and $\tilde{p}_E$ are the restricted MLE's calculated for Farrington-Manning's test statistic. The proposed statistic uses Eldridge et al's (2006) formula to calculate the variance inflation factor (VIF) which is used to adjust the variance of the control group's event rate to account for the correlation among controls in the same matched set. The formula takes into account the possibility of unequal set sizes (i.e. variable-ratio matching, where the number of controls matched to each treated subject is allowed to vary from one matched set to another) by using the coefficient of variation of the set size,

$$VIF_C = 1 + \left(\left[\left\{cv_C^2 \frac{(K-1)}{K} + 1\right\}\overline{n_C}\right] - 1\right) ICC_C\,,$$

where $cv_C = \frac{s_{n_C}}{\overline{n_C}}$ is the coefficient of variation of the set size in the control group. The variance of the set size in the control group is $s_{n_C}^2 = \frac{\sum_{k=1}^{K}(n_{Ck}-\overline{n_C})^2}{(K-1)}$. $\overline{n_C}$ is the mean set size for the control group. n_{Ck} is the size of matched set *k* in the control group. Wu et al (2012) presented several methods for estimating the intra-class correlation coefficient for binary responses. The proposed statistic uses a modified version of Fleiss-Cuzick estimator for the intra-class correlation coefficient (ICC_C) of the control group. The modified Fleiss-Cuzick estimator is based on Farrington-Manning's RMLE of the event rate in the control group instead of the unrestricted MLE. The formula for the ICC_C is

$$ICC_C = 1 - \frac{\sum_{k=1}^{K} \frac{x_{Ck}(n_{Ck} - x_{Ck})}{n_{Ck}}}{(N_C - K)\widetilde{p_C}(1 - \widetilde{p_C})}$$

To account for the possible correlation between subjects in the control group and experimental group who are matched, the proposed statistic uses a modified version of

Obuchowski's method to estimate the covariance between the event rates' of the control group and experimental group. The proposed statistic uses Farrington-Manning's RMLEs ($\tilde{p}_C$ and $\tilde{p}_E$) in estimating the covariance:

$$Cov(\widehat{p_C}, \widehat{p_E}) = \frac{K}{K-1} \sum ((x_{Ck} - n_{Ck} * \tilde{p}_C)(x_{Ek} - \tilde{p}_E))/(N_C N_E)$$

1.3.3 A continuity correction for the adjusted Farrington-Manning test

Correlation between matched subjects can have an impact on the effective sample size. When subjects within a matched set are highly positively correlated, the effective sample size could be very small. As a result, the proposed test statistic (6) will no longer be asymptotically standard normal under the null hypothesis. Therefore, we propose using a modified Yates' continuity correction (CC) for situations in which the sample size is very small or when subjects within a matched set are extremely highly correlated. The continuity correction may improve the normal approximation of the proposed statistic for very small sample sizes or very highly correlated matched subjects. The proposed test statistic with the continuity correction is

$$Z_{AFM_{CC}} = \frac{\hat{p}_C - \hat{p}_E - \delta_0 - CC}{\sqrt{\widehat{\mathrm{Var}}_{AFM}(\hat{p}_C - \hat{p}_E - \delta_0)}} \tag{7}$$

where

$$CC = \frac{1}{2}\left(\frac{1}{N_{Total}^{eff}}\right), \qquad N_{Total}^{eff} = K * (\rho) + \left(N_E + N_c^{eff}\right)(1-\rho),$$

$$N_C^{eff} = \frac{N_C}{VIF_C}, \qquad \rho = \rho(\widehat{p_C}, \widehat{p_E}) = \frac{Cov(\widehat{p_C}, \widehat{p_E})}{\sqrt{\frac{\widetilde{p_C}(1-\widetilde{p_C})}{N_C} * VIF_C}\sqrt{\frac{\widetilde{p_E}(1-\widetilde{p_E})}{N_E}}}$$

1.4. Simulation Study

1.4.1 Parameters and Data-Generation for the Simulation Study

We performed a Monte Carlo simulation study to assess the empirical type I error rate and power of the following test statistics for assessing non-inferiority of an experimental treatment compared with the active control treatment in one-to-many matched studies:

- FM (Farrington-Manning test statistic based on RMLE and assuming independent samples)
- MOM (test statistic for clustered matched-pair data based on methods of moments)
- ICC (intra-class correlation coefficient adjusted test statistic for clustered matched-pair data based on RMLE)
- O (Obuchowski-based test statistic for clustered matched-pair data)
- GEE Indep (GEE-based test statistic using an independent working correlation structure)
- GEE Exch (GEE-based test statistic using an exchangeable working correlation structure)

- GEE Un (GEE-based test statistic using an unstructured working correlation structure)
- AFM (adjusted Farrington-Manning statistic for one-to-many matched data)
- AFM CC (continuity-corrected adjusted Farrington-Manning statistic)

Pre-specified parameters in the simulation study are the number of clusters or matched sets (*K*), the number of controls matched to an experimental treated subject in the *k*th matched set (n_k), the distribution or variability of the size of the matched sets, the probabilities of having an event for the control and experimental procedures (p_C and p_E, respectively) and the non-inferiority margin (δ_0), which is the materially unacceptable difference between the control and experimental procedures. We specify two within-matched set correlations: the correlation of the responses in the control group (r_C) and the correlation of the responses between the experimental treated subject and the control subjects (r_{CE}). The simulation study was conducted using SAS version 9.3 via PROC IML and macro-language. SAS PROC GENMOD was used to run GEE models. We use a data-generating process similar to the one used by Nam et al (2009). For each matched set, we generate a random vector from a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance $\mathbf{\Sigma}$ where correlation coefficients (r_C and r_{EC}) are specified. Denote matched set normal variates as *z*. Then, we generate a binomial response for each procedure. Define *y* as *y*=1 if z ≤ c or *y* = 0 otherwise, where *c* is a cut-off point which satisfies Pr(z ≤ c) = p_C if z is a normal variate from the control group and Pr(z ≤ c) = p_E if z is a normal variate from the experimental group. When the size of the matched sets (n_k) varies, we consider two kinds of distributions: the value of n_k is generated from a uniform distribution or from a beta(2,3) distribution. In our simulation study, 50,000 data sets were generated for each configuration to calculate empirical Type I error rates and power. If the expected Type I error rate of a test is indeed a nominal 0.025 level, then a 95% confidence interval for α=0.025 is (0.024, 0.026) from 50,000 simulated data. Those empirical Type I error rates above the interval are shown with a caret sign (^) to indicate inflated Type I error rates and those below the interval are shown with an asterisk sign (*) to indicate a conservative test. If the GEE models failed to converge for any of the 50,000 data sets that were generated for each scenario, then we report in brackets ([ ]) the percentage of the 50,000 data sets that the GEE model successfully converged.

Tables 1.1-1.105 summarize simulated Type I error rates and power of tests for K = (200, 500, 1000), $n_k \leq (1, 2, 5)$ assuming either a constant, uniform, or non-uniform (beta) distribution, p_C = (0.05, 0.1, 0.2, 0.3, 0.5), δ_0 = (-0.02, -0.05, -0.1), and various combinations of r_C = (0, 0.2, 0.4, 0.6, 0.8) and r_{CE} = (0, 0.2, 0.4, 0.5, 0.6, 0.8). Empirical type I error rates were calculated under the null hypothesis (p_E = p_C - δ_0) and power was calculated under the alternative hypothesis (p_E = p_C).

1.4.2. Simulation Results for 1-to-1 Matched Designs

Tables 1.1-1.15 display empirical Type I error rates and power of tests for 1:1 matched designs ($n_k = 1$) under different scenarios. In the setting of 1:1 matching, the empirical Type I error rate for the *FM* test statistic stays below the nominal level of 2.5% and decreases as the correlation (r_{CE}) increases, but *FM* also becomes substantially less powerful than the other test statistics as the correlation (r_{CE}) increases. In the 1:1

matching setting, *ICC* stays satisfactorily close to the nominal level of 2.5% for all scenarios considered. *MOM*, *O*, and *GEE*-based statistics has similar empirical Type I error rates and are inflated for relatively low event rates ($p_C < 0.1$). *AFM* maintained satisfactorily close to nominal level of 2.5% except in the case for very high correlation (r_{CE}=0.8). In the cases of high correlation, the Type I error rate for the continuity-corrected *AFM* statistic ($AFM\ CC$) was significantly close to or below the nominal level of 2.5%.

1.4.3 Simulation Results for 1-to-Many Matched Designs

Tables 1.16-1.30 display empirical Type I error rates and power of tests for 1-to-2 matched designs where each treated subject is matched to two control subjects (i.e. $n_k = 2$ for all matched sets *k*). Table 1.31-1.45 displays empirical Type I error rates and power of tests for 1-to-≤2 matched designs where each treated subject is matched to up to two control subjects (i.e. $n_k \leq 2$) and is uniformly distributed across matched sets. Tables 1.46-1.60 displays empirical Type I error rates and power of tests, for 1-to-≤2 matched designs, where each treated subject is matched to up to two control subjects (i.e. $n_k \leq 2$) and is non-uniformly distributed across matched sets. Tables 1.61-1.75 display empirical Type I error rates and power of tests for 1-to-5 matched designs where each treated subject is matched to five control subjects (i.e. $n_k = 5$ for all matched sets *k*). Table 1.76-1.90 displays empirical Type I error rates and power of tests for 1-to-≤5 matched designs where each treated subject is matched to up to five control subjects (i.e. $n_k \leq 5$) and is uniformly distributed across matched sets. Tables 1.91-1.105 displays empirical Type I error rates and power of tests, for 1-to-≤5 matched designs, where each treated subject is matched to up to five control subjects (i.e. $n_k \leq 5$) and is non-uniformly distributed.

For one-to-many matched designs, empirical Type I error rates for *FM* test decrease and stay significantly below the nominal level as the correlation of the responses between treated subjects and matched controls (r_{CE}) increases, but *FM* also becomes substantially less powerful than the other test statistics as the correlation (r_{CE}) increases. However, in cases where the responses between treated subjects and matched controls are independent (r_{CE}=0), empirical Type I error rates for *FM* increases and is above the nominal level as the correlation of the responses in the control group (r_C) increases.

Empirical Type I error rates for $MOM, ICC, O,$ and GEE-based test statistics are inflated for various correlation structures in one-to-many matched designs, with Type I error rates becoming less inflated as the event rate (p_C) increases. The *GEE*-based test statistic that uses an unstructured working correlation structure (*GEE Un*) fails to converge in some scenarios where the number of controls matched to each treated subject is allowed to vary from one matched set to another, or when the number of controls matched to each treated subject is relatively high ($n_k = 5$). We also tried to assume a Poisson distribution for the outcome in the GEE model, but the results were similar to those that assume a binomial distribution for the outcome.

Except in cases of high correlation, Type I error rates for the new proposed test (*AFM*) are satisfactorily close to or slightly below the nominal level of 2.5% in one-to-many matched study designs (regardless of the distribution of the number of controls matched to each treated subject). Type I error rates for the continuity-corrected *AFM* test statistic (*AFM CC*) is satisfactorily close to or slightly below the nominal level of 2.5% for all different correlation structures that were considered (including those with high within-matched set correlation) in one-to-many matched study designs. For all scenarios considered, *AFM CC* is always more conservative than *AFM* and $AFM\ CC$ is slightly less powerful than AFM.

4.3.1. Operating characteristics corresponding to different number of matched sets

Figure 1A and Figure 1B display the empirical Type I error rates and power, respectively, as the number of matched sets changes for the following scenario: $n_k \leq 5$, n_k~beta(2,3), p_C=0.1, δ_0 = -0.05, r_C=0.4, r_{CE}=0.4. In this scenario, Type I error rates for *MOM, ICC, O, GEE Ind, and GEE Exch* are significantly above the nominal level of 2.5% for K=(200, 500, 1000), with Type I error rates for *MOM, ICC, O, GEE Ind,* and *GEE Exch* decreasing as the number of matched sets increases. In contrast, Type I error rates for *FM, AFM,* and *AFM CC* are significantly close to or below the nominal level of 2.5%, with *FM* being the most conservative but also the least powerful. Power for all test statistics increases as the number of matched sets increases.

Figure 1A. Empirical Type I Error Rate vs. Number of Matched Sets: nk≤5, nk~beta(2,3), pC=0.1, δ0 = -0.05, rC=0.4, rCE=0.4

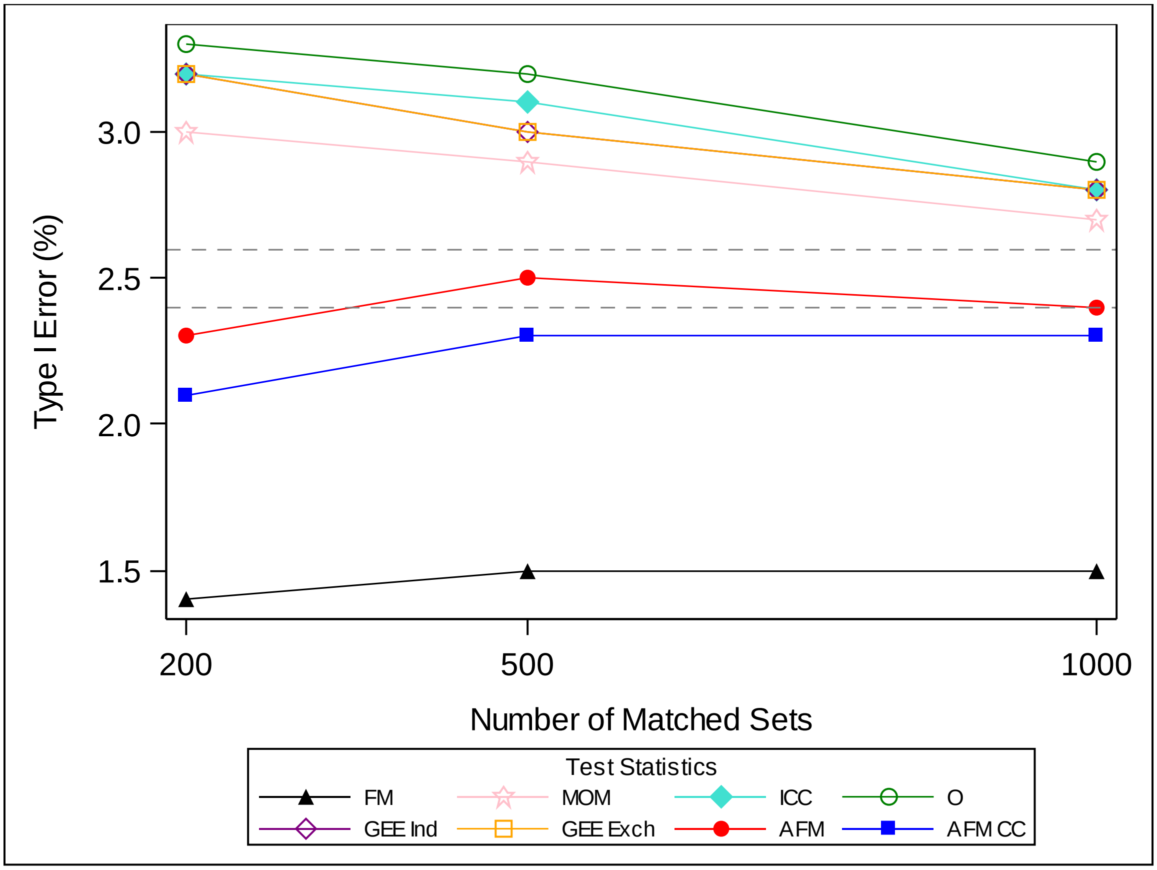
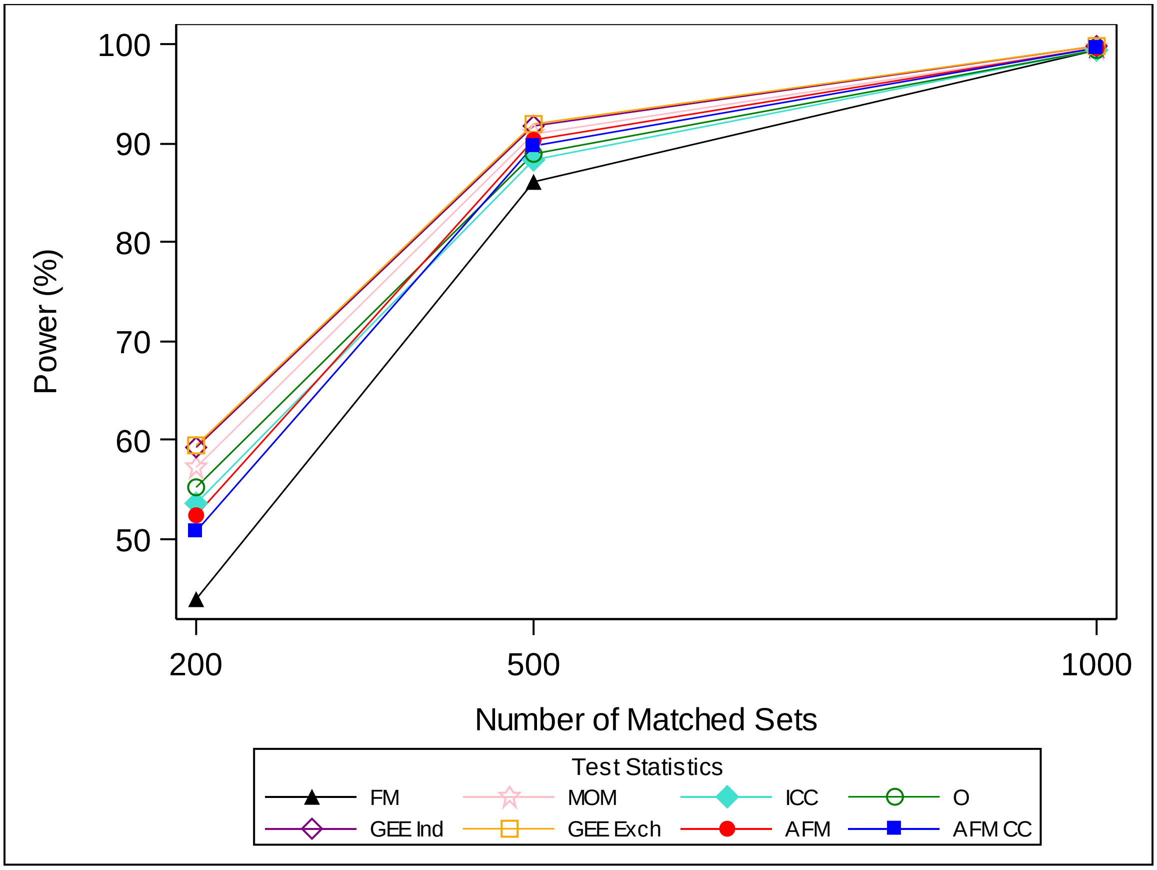


Figure 1B. Empirical Power vs. Number of Matched Sets: $n_k \leq 5$, n_k~beta(2,3), p_C=0.1, δ_0 = -0.05, r_C=0.4, r_{CE}=0.4



4.3.2. Operating characteristics corresponding to different number of controls matched to each treated subject

Figure 2A and Figure 2B display the empirical Type I error rates and power, respectively, as the number of controls matched to each treated subject changes for the following scenario: k=500, n_k~constant, p_C=0.1, δ_0 = -0.05, r_C=0.4, r_{CE}=0.4. In this scenario, Type I error rates for *MOM, O, GEE Ind, and GEE Exch* are significantly above the nominal level of 2.5% for n_k = (1, 2, 5). Type I error rates for *ICC* is close to the nominal level for n_k = 1, but is inflated for n_k = (2, 5). In contrast, Type I error rates for

FM, AFM, and *AFM CC* are significantly close to or below the nominal level of 2.5%, with *FM* being the most conservative but also the least powerful. Power for all tests increases as the number of controls matched to each treated subject increases.

Figure 2A. Empirical Type I Error Rate vs. Number of Controls Matched to Each Treated Subject: k=500, n_k~constant, p_C=0.1, δ_0 = -0.05, r_C=0.4, r_{CE}=0.4

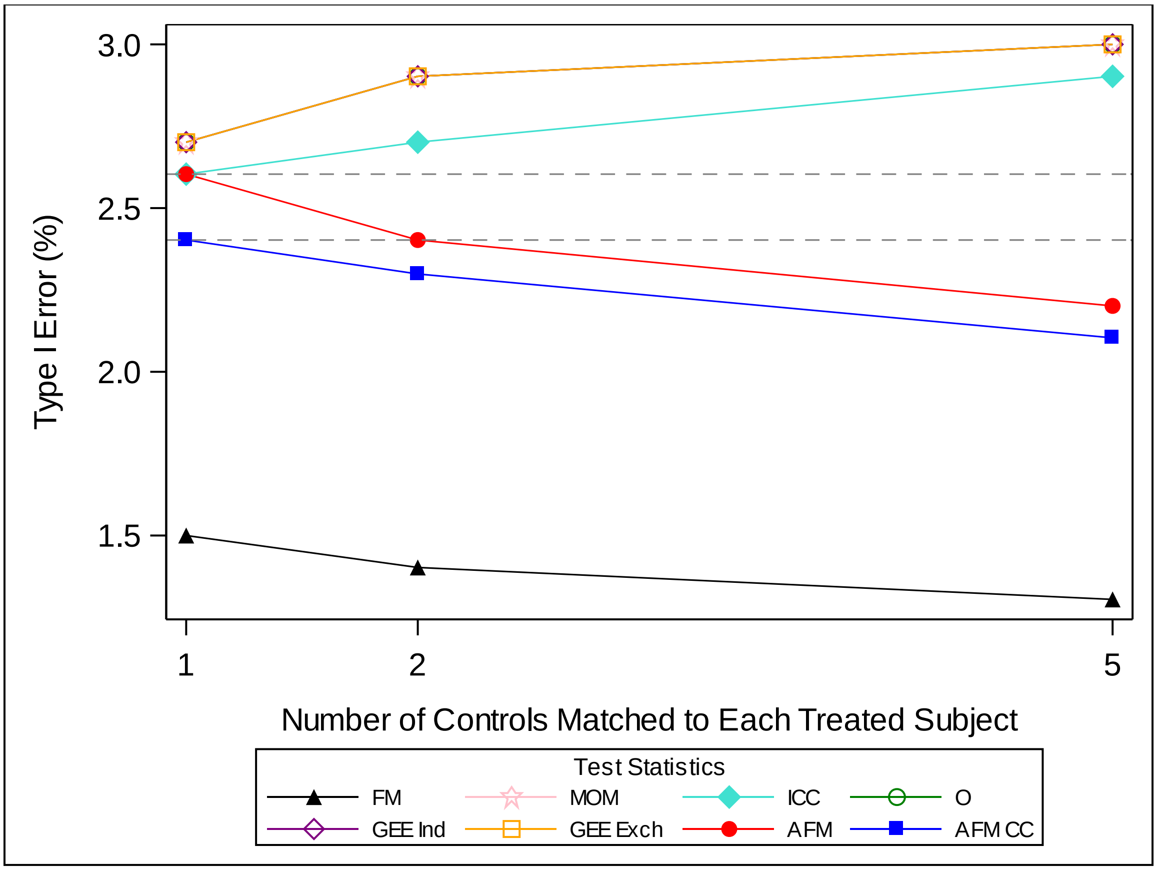
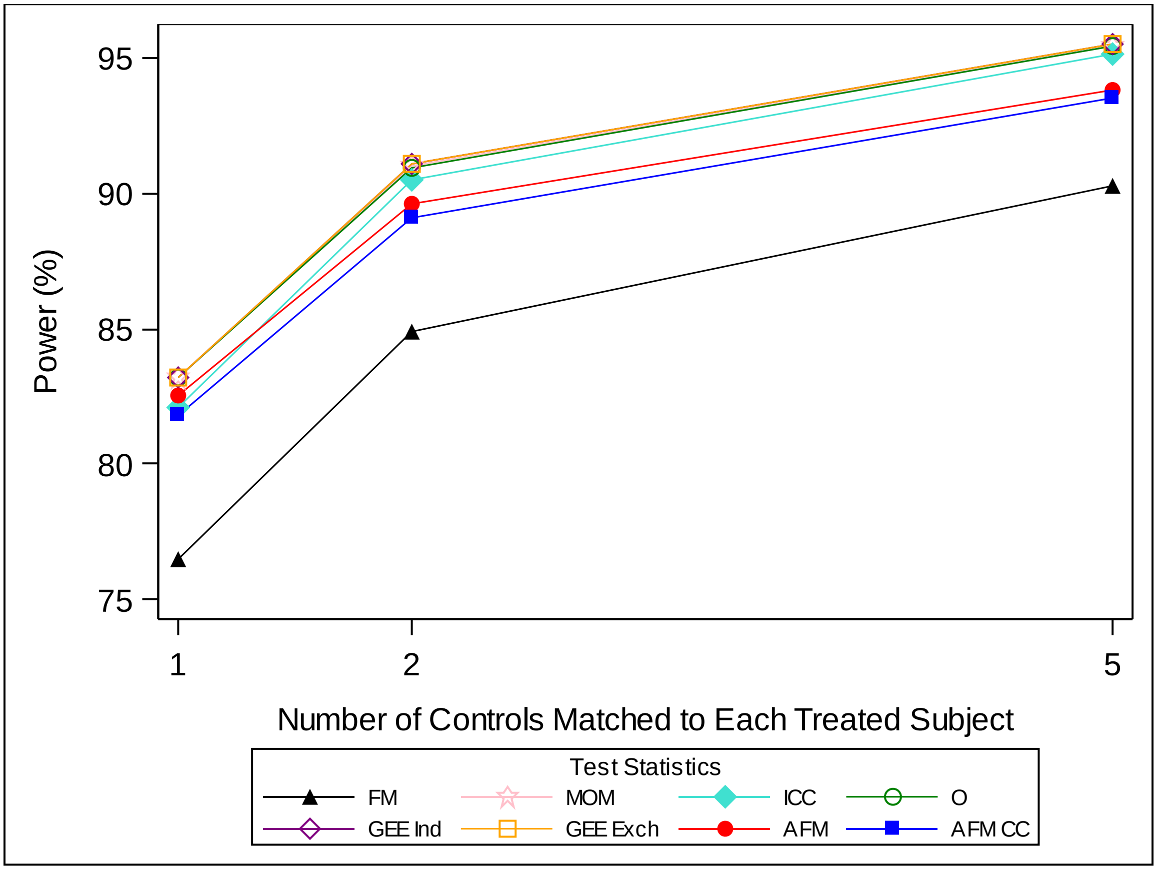


Figure 2B. Empirical Power vs. Number of Controls Matched to Each Treated Subject: k=500, n_k~constant, p_C=0.1, δ_0 = -0.05, r_C=0.4, r_{CE}=0.4



4.3.3. Operating characteristics corresponding to different distributions of the number of controls matched to each treated subject

Figure 3A and Figure 3B display the empirical Type I error rates and power, respectively, as the distribution of the number of controls matched to each treated subject changes for the following scenario: k=500, $n_k \leq 5$, p_C=0.1, δ_0 = -0.05, r_C=0.4, r_{CE}=0.4. In this scenario, Type I error rates for *MOM, ICC, O, GEE Ind, and GEE Exch* are significantly above the nominal level of 2.5% for non-uniform, uniform, or constant distributions. In contrast, Type I error rates for *FM, AFM,* and *AFM CC* are significantly close to or below the nominal level of 2.5%, with *FM* being the most conservative but also the least powerful.

Figure 3A. Empirical Type I Error Rate vs. Distribution of the Number of Controls Matched to Each Treated Subject: k=500, n_k≤5, p_C=0.1, δ_0 = -0.05, r_C=0.4, r_{CE}=0.4

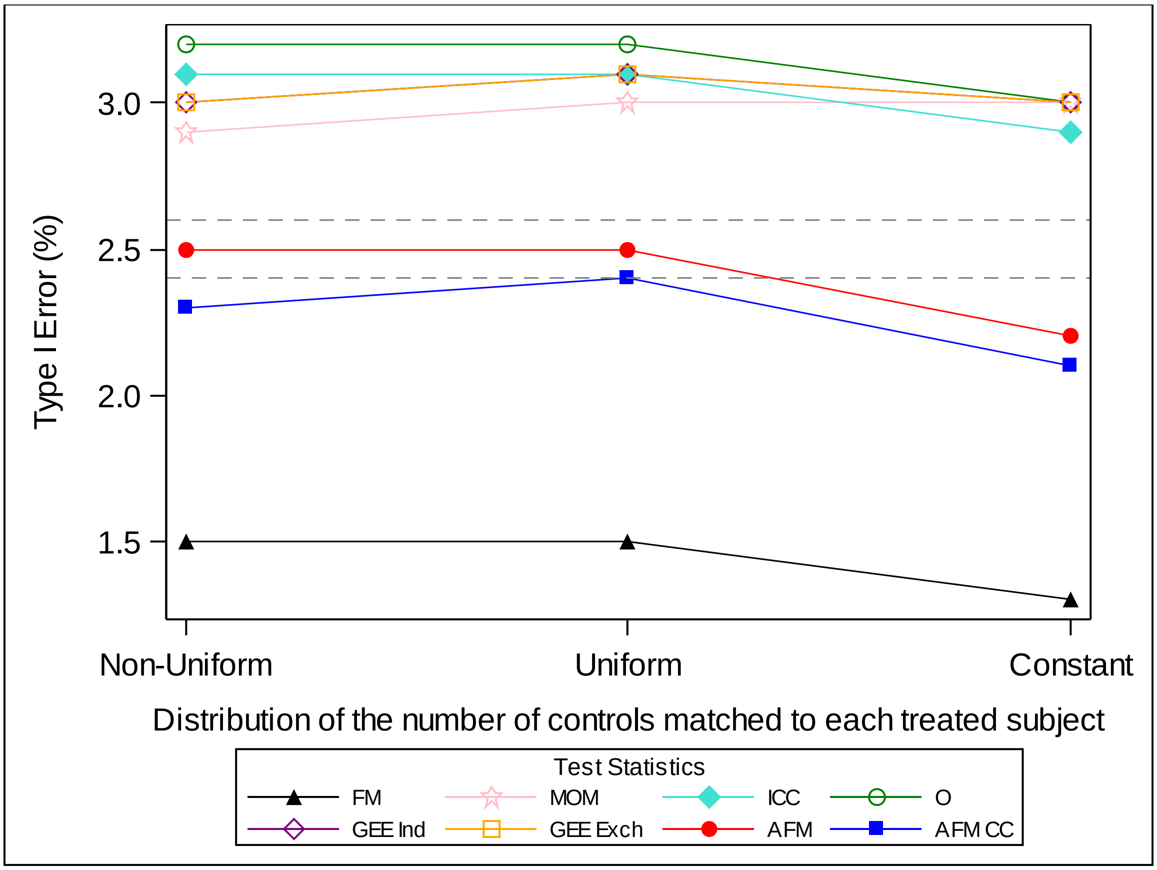
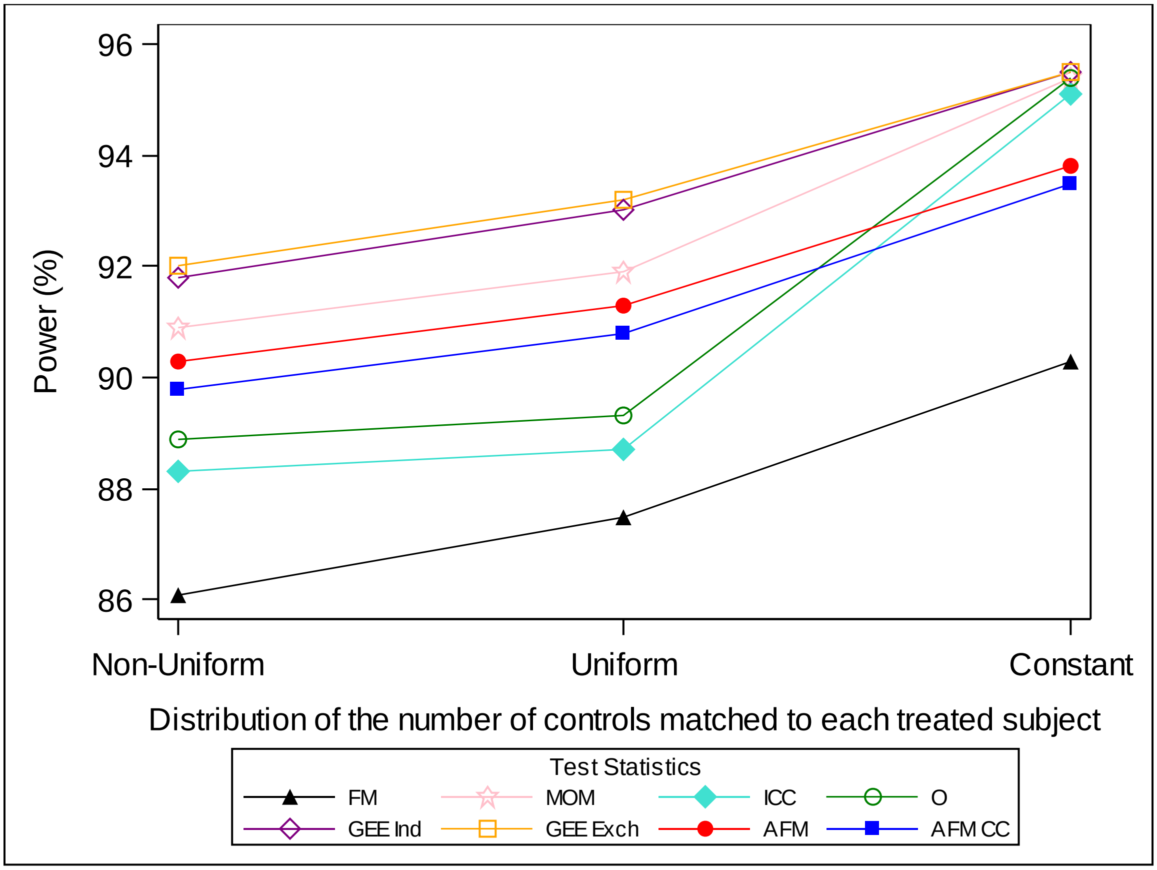


Figure 3B. Empirical Power vs. Distribution of the Number of Controls Matched to Each Treated Subject: k=500, $n_k \le 5$, p_C=0.1, δ_0 = -0.05, r_C=0.4, r_{CE}=0.4



4.3.4. Operating characteristics corresponding to different correlations of the responses within matched sets

Figure 4A and Figure 4B display the empirical Type I error rates and power, respectively, as the correlation of the responses within matched sets (r_C and r_{CE}) changes for the following scenario: k=500, n_k ~ beta(2,3), p_c=0.1, δ_0 = -0.05, and $r_c = r_{CE}$. In this scenario, Type I error rates for *MOM, ICC, O, GEE Ind, and GEE Exch* are significantly above the nominal level of 2.5% for (r_C, r_{CE}) = {(0,0), (0.2, 0.2), (0.4, 0.4), (0.6, 0.6), (0.8, 0.8)}. In contrast, Type I error rates for *FM, AFM,* and *AFM CC* are significantly close to or below the nominal level of 2.5%, with *FM* being the most conservative but also the least powerful. The Type I error rate for *FM* drastically decreases as the correlation of the responses within matched sets increases, but also becomes less powerful compared to *AFM* and *AFM CC*. In contrast, the *AFM* and *AFM CC* always stay significantly close to or slightly below the nominal level of 2.5% as the correlation increases. Under this scenario, when both r_C and r_{CE} increase at the same time and r_C=r_{CE}, empirical power also increases. However, as seen in Tables 1.16-1.105, when r_{CE} is held constant, empirical power decreases as r_C increases for all test statistics that account for correlated matched data.

Figure 4A. Empirical Type I Error Rate vs. Correlation within Matched Sets: k=500, n_k≤5, n_k~beta(2,3), p_C=0.1, δ_0 = -0.05, r_C=r_{CE}

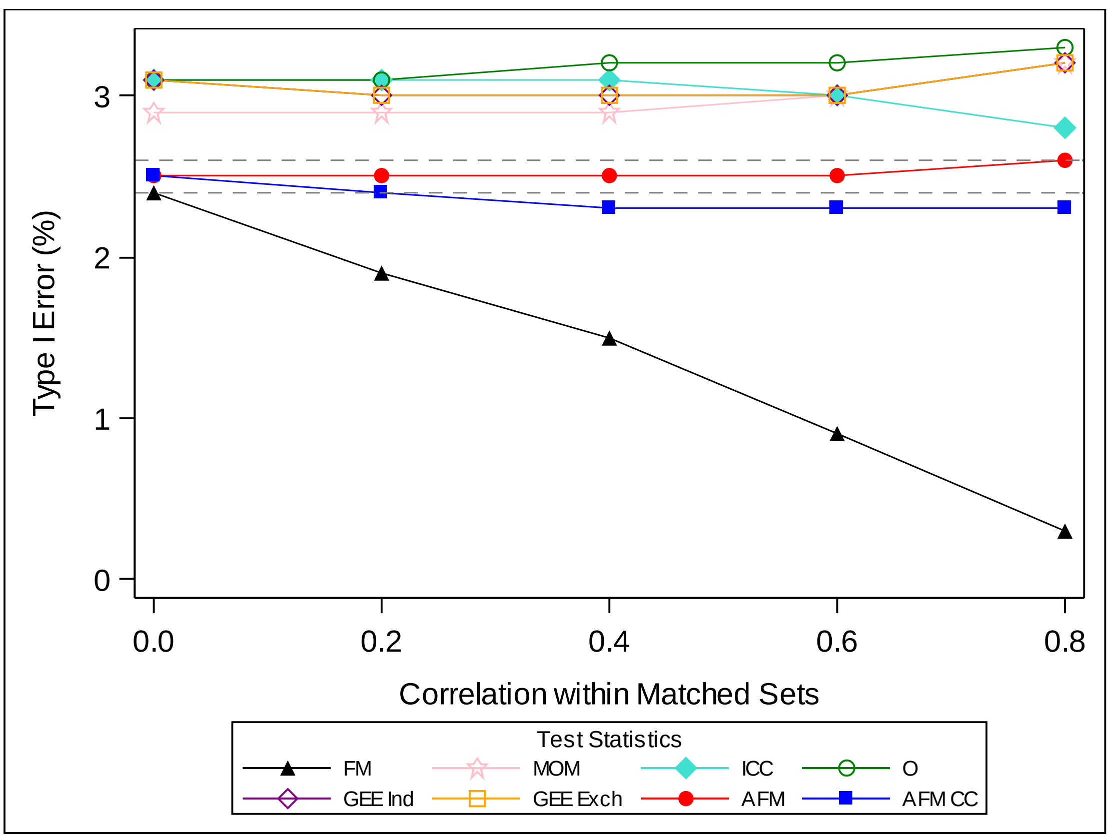
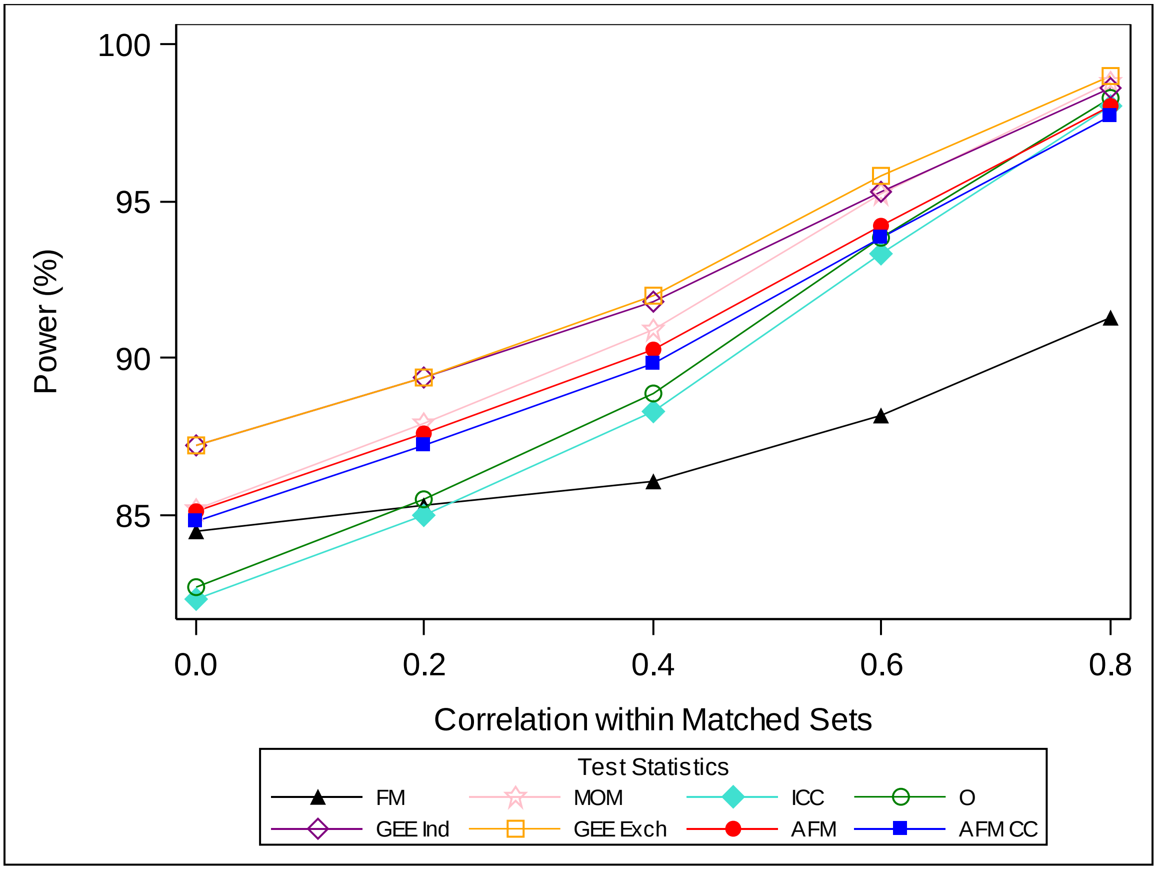


Figure 4B. Empirical Power vs. Correlation within Matched Sets: k=500, n_k≤5, n_k~beta(2,3), p_C=0.1, δ_0 = -0.05, r_C=r_{CE}



1.5. Discussion

Farrington-Manning (1990) developed a test statistic for non-inferiority assuming treatment groups are independent. Nam (1997) developed a test for non-inferiority for 1:1 matched pairs based on RMLE. Several statistics were developed to test non-inferiority for clustered 1:1 matched pair data: Durkalski et al (2003) presented a test statistic based on method of moments; Nam and Kwon (2009) presented an ICC-adjusted statistic and a modification of Obuchowski's method. We investigated these tests and GEE-based test statistics to see how they perform when assessing non-inferiority in one-to-many matched trials. We also proposed a new test statistic that was specifically developed to analyze one-to-many matched data. This new statistic was based on adjusting the Farrington-Manning statistic using the idea of variance inflation factor and Obuchowski's estimate of the covariance. We also proposed a continuity correction for the proposed test for cases with high within-matched set correlation.

The empirical type I error rate for *FM* statistic is not stable across different correlation structures. The empirical Type I error rate for the ICC-adjusted statistic (*ICC*) stays close to the nominal level when analyzing 1:1 matched data. However, when more than one control is matched to an experimental (1: ≤M), the empirical Type I error rate for *ICC, MO, O,* and *GEE*-based test statistics are inflated for numerous correlation structures. In contrast, the empirical Type I error rate of the newly proposed adjusted Farrington-Manning statistic (*AFM*) stays close to or slightly below the nominal level for most scenarios. For situations in which subjects within a matched set are extremely highly correlated, the continuity-corrected statistic *AFM CC* stays close to the nominal level with little cost in power relative to the *AFM* statistic. We suggest that the new non-inferiority statistic (*AFM*) and/or its continuity-correction (*AFM CC*) will control Type I error at or below the nominal level (2.5%) and are more powerful than *FM* for studies where each experimental treated subject is matched to ≥ 1 control subject.

2. COMPARING A METHOD FOR INDEPENDENT DATA TO A METHOD FOR MATCHED DATA WHEN ASSESSING NON-INFERIORITY VIA RISK DIFFERENCE IN PROPENSITY-SCORE MATCHED STUDIES WHEN THERE ARE MORE CONTROL SUBJECTS THAN EXPERIMENTAL SUBJECTS

2.1. Introduction

Given the large number of proven effective treatments in some therapeutic areas, non-inferiority safety trials are increasingly being used in clinical research to evaluate the post-market safety of an experimental treatment. The primary objective of a non-inferiority safety trial is to demonstrate that an experimental treatment is not materially worse than a comparator by a pre-specified amount (i.e. non-inferiority margin) with respect to a safety outcome (e.g. major adverse cardiovascular event). In other words, non-inferiority safety trials are conducted to show that a new treatment will not result in a clinically unacceptable increase in risk compare to a control treatment.

Randomized controlled trials (RCTs) are considered the gold standard for assessing non-inferiority of an experimental treatment compared to a control treatment. However, for ethical or practical reasons, it is sometimes not possible to conduct RCTs. Therefore, non-randomized (observational) studies are commonly used to address safety questions for medical drugs, biologics, and devices in a regulatory setting (Levenson & Yue, 2013). However, in non-randomized trials, there is a likelihood of imbalance between treatment groups on clinically relevant (observed and unobserved) baseline characteristics because treatment selection is often influenced by subject characteristics. This could potentially lead to biased estimates of the treatment effect.

Propensity score methodology is commonly used in non-randomized studies to account for differences in subjects' baseline characteristics. A propensity score is defined as the probability of receiving the experimental treatment (versus control) conditioned on the individual's observed baseline characteristics (Rosenbaum & Rubin, 1983). When conditioning on the propensity score, the distribution of observed baseline covariates will be similar between the experimental and control groups, and thus allowing one to obtain an unbiased estimate of the treatment effect (Rosenbaum & Rubin, 1983). Propensity-score methodology is increasingly being used in clinical research. For example, propensity-score methods have been used in premarket nonrandomized medical device studies with concurrent control, device postmarket observational studies, and pharmacoepidemiologic safety studies using electronic healthcare data (Levenson & Yue, 2013). The Food and Drug Administration (FDA) has suggested the use of propensity score methodology as one possible approach to address confounding in pharmacoepidemiologic safety studies (Food and Drug Administration, 2013).

Matching on propensity score is a common technique used to reduce the imbalance between treatment groups on clinically relevant baseline covariates and to make unbiased inferences about treatment effects in non-randomized trials. Propensity-score matching is frequently used in cardiovascular research studies (Deb, et al., 2016; Ko, et al., 2008; Boening, et al., 2003; Ko, et al., 2009). It also been used in some non-randomized non-inferiority studies (Pyo, Lee, Min, Lee, & et al., 2016; Kereiakes, Yeh, Massaro, Driscoll-Shempp, & et al., 2015). In propensity-score matching, a matched sample is formed consisting of sets of subjects from the experimental and control groups with similar values of the propensity score (Rosenbaum & Rubin, 1985). Then, one directly carries out treatment comparisons on the primary outcome in the matched sample (Austin, An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies, 2011). However, there is a lack of consensus as to the appropriate statistical method to use when assessing the statistical significance of the treatment effect in a propensity-score matched sample. Some believe that the treatment comparison on the outcome in the matched sample should be performed as if the sample had been generated through randomization (Stuart, 2010), without accounting for the matching in the analysis. That is, they believe researchers can do the exact same analysis they would have done using the data from a RCTs, but using the matched data instead (Ho, Imai, King, & Stuart, 2007) since propensity score matching allows one to mimic some of the particular characteristics of a RCT, such as the distribution of observed baseline covariates being similar between treatment groups. Schafer and Kang (2008) suggest that matched subjects should be regarded as independent because there is no reason to believe that the outcomes of propensity-matched individuals are correlated in any way, and thus it is not necessary to account for the matching. They argue that the theory behind propensity score does not guarantee that matched individuals will have the same values on the full set of covariates – in fact, two individuals with the same propensity score may have very different values of the covariates. The theory of propensity scores only says that *groups* of individuals with similar propensity scores will have similar covariate *distributions* (Stuart, 2008). A critical evaluation of 47 articles that were published between 1996 and 2003 in the medical literature and that employed propensity-score matching found that 34 of the 47 articles used statistical methods that did not account for the matched nature of the data (Austin, A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003, 2008).

On the other hand, some in the literature suggest that a propensity-score matched sample does not consist of independent observations, and therefore any statistical analysis must account for the matched nature of the sample (Austin, Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples, 2011). More specifically, the variance should be calculated using a method accounting for correlated matched data (Imbens, 2004). They argue that, on average, there are no systematic differences in baseline covariates between the experimental and control groups in the propensity-score matched sample. However, in the overall (unmatched) sample, systematic differences usually exist between the experimental and control groups in non-randomized studies. This implies that matched subjects are, on average, more similar with respect to baseline covariates than are randomly selected subjects (Austin, A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003, 2008). When confounding is present, baseline covariates are related to the outcome. Therefore, the outcomes of matched subjects are likely to be more similar to one another compared to the outcomes of randomly selected subjects (Austin, Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples, 2011). Hence, the experimental and control groups in the matched sample do not form two independent samples, and therefore, all assessments of statistical significance must account for the lack of independence in the propensity-score matched sample (Austin, A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003, 2008).

To address this statistical debate, Austin (Austin, Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples, 2011) performed a Monte Carlo simulation study comparing statistical methods for matched data (i.e. McNemar's test) and independent data (i.e. Pearson Chi-Squared test) when conducting a two-sided hypothesis test of the risk difference in propensity-score *one-to-one* matched samples. The simulation study demonstrated that for a range of scenarios, variance estimators that account for matching more accurately reflected the sampling variability of the estimated treatment effect. From the study, Austin also found that the use of methods for matched data resulted in empirical type I error rates that were closer to the advertised rate (i.e. 0.05 significance level), with statistical methods for independent samples being more conservative. Therefore, Austin recommended using statistical methods that account for the matched nature of the data when using one-to-one propensity-score matched samples to conduct two-sided statistical tests about the risk difference. However, there are a few shortcomings in Austin's simulation study. The study did not compare the statistical methods with respect to statistical power. Also, the simulation study assumed that the effect sizes of the baseline covariates on the probability of being treated and on the probability of having the event of interest to be the same, which is rarely the case in the real world. The study also did not address the impact of *one-to-many* matching on the operating characteristics of non-inferiority statistical methods.

The most common (and simplest) approach to propensity-score matching is one-to-one matching, in which pairs of experimental and control subjects with similar propensity score values are formed. A criticism of one-to-one matching on the propensity score is that it is "wasteful" of sample size, and thus has lower statistical power. Therefore, when there are large numbers of control subjects, some studies have used one-to-many matching on the propensity-score (Boening, et al., 2003). Selecting multiple controls for each treated individual will generally increase bias in the estimated treatment effect, but utilizing multiple matches can increase precision and power due to the larger matched sample size (Stuart, 2010). One way to reduce the bias from one-to-many matching is by allowing the number of controls matched to each experimental subject to vary (i.e. variable ratio matching), rather than having the number of controls matched to each experimental subject to be fixed (i.e. fixed ratio matching) (Ming & Rosenbaum, 2000). A Monte Carlo simulation study by Rassen et al. (Rassen, et al., 2012) has shown that variable ratio matching outperformed fixed ratio matching with respect to bias, precision, and mean squared error (MSE); and that one-to-many variable ratio matching yielded higher precision than one-to-one matching at the cost of a small increase in bias. Another Monte Carlo simulation study by Austin (Austin, Statistical Criteria for Selecting the Optimal Number of Untreated Subjects Matched to Each Treated Subject When Using Many-to-One Matching on the Propensity Score, 2010) has shown, for many scenarios, that MSE is minimized when matching either 1 or 2 controls to each treated subject when using propensity-score matching. The previously mentioned Monte Carlo simulation studies focused on continuous outcomes. However, there is also a need to investigate the impact of one-to-many matching when the outcome of interest is dichotomous, which is a common outcome in medical and clinical research. The aforementioned simulation studies examined the effect of increasing the number of controls subjects matched to each treated subject and/or the type of matching (i.e. variable ratio matching vs. fixed ratio matching) on bias, precision, and MSE; but the studies did not address how one-to-many matching impacts the operating characteristics (i.e. type I error and power) of inferential tests, in particular, non-inferiority tests. One reason for this is that one-to-many matching can make estimation of the variance of the estimated treatment effect more difficult, especially if one wants to account for the matching (Austin, Statistical Criteria for Selecting the Optimal Number of Untreated Subjects Matched to Each Treated Subject When Using Many-to-One Matching on the Propensity Score, 2010). Therefore, when multiple controls are matched to each experimental subject, it is unclear how the statistical significance of the treatment effect (e.g. risk difference) should be determined; especially when assessing non-inferiority (Austin, Statistical Criteria for Selecting the Optimal Number of Untreated Subjects Matched to Each Treated Subject When Using Many-to-One Matching on the Propensity Score, 2010). As a result, there is a lack of methodological research addressing what is the appropriate statistical method to use when assessing non-inferiority in a *one-to-many* propensity-score matched study.

In order to address and expand the debate regarding whether or not statistical analyses should account for the matched nature of the data in propensity-score matched studies, we compare a statistical method that assumes independent samples (i.e. Farrington-Manning test) to a new statistical method proposed in chapter one of this dissertation (i.e. adjusted Farrington-Manning test) that assumes correlated matched samples for assessing non-inferiority via risk difference in non-randomized trials using *one-to-many* matching on the propensity score. Under various propensity-score matching scenarios, we compare the empirical type I error rate and power of Farrington-Manning's test (FM) to the adjusted Farrington-Manning's test (AFM). We also compare the empirical coverage rates and width of estimated 95% confidence intervals when standard errors are estimated using methods for matched data (i.e. AFM approach) compared with when methods for independent samples are used (i.e. original FM approach). We compare the bias of the estimated treatment effect when assuming independent data versus when assuming one-to-many matched data. Finally, we also compare the standard deviation of the empirical sampling distribution of the risk difference with the estimated standard error of the risk difference that is used to calculate the FM test statistic and the AFM test statistic. These objectives will be addressed using Monte Carlo simulations. In section 2, we present underlying theory of propensity score matching. In section 3, we describe the methods used in the Monte Carlo simulation study. In section 4, we present the results of the Monte Carlo simulation study. In section 5, we present a numerical example. In section 6, we summarize our findings and provide a discussion.

2.2. The Rubin Causal Model Potential Outcomes Framework

Rubin's potential outcomes framework (Rubin, 1974) provides the underlying theory of propensity-score matching. In this section, we will briefly describe Rubin's potential outcomes framework. Let T_i denote a binary treatment for individual i (T_i = 1 denoting experimental treatment; T_i = 0 denoting control treatment). Each individual has a *pair* of potential outcomes. Let $Y_i(T_i = 1) = Y_i(1)$ denote the potential outcome that would be observed if individual i receives the experimental treatment, and let $Y_i(T_i = 0) = Y_i(0)$ denote the potential outcome that would be observed if individual i receives the control treatment. Causal effects are comparisons of these *pair* of potential outcomes. The causal effect of treatment for individual i is $\Delta_i = Y_i(0)$ - $Y_i(1)$. The average treatment effect (ATE) is the average effect of moving the entire population from control to experimental treatment: ATE = E[Y(0) – Y(1)]. The average treatment effect for the treated (ATT) is the average effect of moving the population of experimental treated subjects from experimental to control treatment: ATT = E[Y(0)-Y(1)|T = 1] (Imbens, 2004). The fundamental problem of causal inference is that each person gets either the experimental treatment or control, so only one of the potential outcomes is observed for each individual. Estimating causal effects involves estimating the unobserved potential outcomes. In an RCT, estimates of the ATE and ATT coincide because, due to randomization, the experimental treated population will not, on average, differ systematically from the overall population (Austin, An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies, 2011). Generally, propensity-score matching nearly always estimates the ATT, rather than ATE

(Imbens, 2004), because it matches control individuals to the experimental treated group and discards controls who are not selected as matches (Stuart, Matching Methods for Causal Inference: A Review and a Look Forward, 2010). The remaining (i.e. matched) controls will have the same (or nearly the same) distributions of baseline covariates as the experimental treated group. Therefore, the observed outcomes of matched controls are used to estimate the unobserved potential outcomes that would be expressed if experimental treated subjects receive the control treatment, thus allowing one to estimate the ATT.

2.3. Monte Carlo Simulation Study - Methods

In this section, we describe the Monte Carlo simulation study that was conducted to compare a method that assumes independent samples (e.g. FM test statistic) to a newly proposed method (e.g. AFM test statistic) that assumes correlated matched data when assessing non-inferiority via risk differences in one-to-many propensity-score matched studies.

2.3.1. Data-generating process

We use a data-generating process similar to the one used in a prior study that compared statistical methods for matched vs. independent data when making inferences (two-sided tests) about risk differences in one-to-one propensity-score matched samples (Austin, Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples, 2011). First, we randomly generated a data set for a specified number of subjects. In this simulation study, we examined scenarios with 2,000 and 20,000 subjects. For each subject, we randomly generated 6 baseline covariates ($X_1 - X_6$) under a specified covariate scenario. We examined six different covariate scenarios. In the first covariate scenario, we generated three independent standard normal random variables ($X_1 - X_3$) and three Bernoulli random variables ($X_4 - X_6$). In the second covariate scenario, we generated three standard normal random variables ($X_1 - X_3$) and three Bernoulli random variables ($X_4 - X_6$) that were correlated. In the third covariate scenario, we generated six independent standard normal covariates. In the fourth covariate scenario, we generated six correlated standard normal covariates. In the fifth covariate scenario, we generated six independent Bernoulli random variables. In the sixth covariate scenario, we generated six correlated Bernoulli random variables. The parameter value for all Bernoulli random variables was 0.5. When generating correlated random variables, we first generated the variables from a multivariate normal distribution such that the mean and variance of each random variable were equal to 0 and 1, respectively, while the correlation between pairs of random variables was equal to 0.25. Denote these correlated normal variates as *z*. Then, we generate correlated Bernoulli (p=0.5) variates *y*, by defining *y* as y=1 if z ≤ 0 or y=0 otherwise.

After generating the six baseline covariates for each subject *i* ($X_{1i} - X_{6i}$), we then calculate the subject-specific probability of receiving the experimental treatment ($p_{i,treat}$) using the following treatment-selection logistic model under a specified scenario:

$$\text{logit}(p_{i,treat}) = \alpha_{0,treat} + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_1 X_{4i} + \alpha_2 X_{5i} + \alpha_3 X_{6i}$$

We examined 3 different treatment-selection logistic model scenarios. Under a weak treatment-selection model, the coefficients for the baseline covariates in the above model were as follows: $\alpha_1 = \log(1.1)$, $\alpha_2 = \log(1.25)$, and $\alpha_3 = \log(1.5)$. Under a strong treatment-selection model, the coefficients for the baseline covariates in the above model were as follows: $\alpha_1 = \log(1.75)$, $\alpha_2 = \log(2)$, and $\alpha_3 = \log(2.5)$. Under a mixed treatment-selection model, the coefficients for the baseline covariates in the above model were as follows: $\alpha_1 = \log(1.1)$, $\alpha_2 = \log(2)$, and $\alpha_3 = \log(1.5)$. The intercept $\alpha_{0,treat}$ was chosen so that approximately 240 of the subjects would be exposed to the experimental treatment (e.g. 1.2% of the 20,000 simulated subjects would be treated with the experimental treatment). The reader is referred to Appendix A.1 of this chapter for a more detailed explanation of how the value of $\alpha_{0,treat}$ was determined. We then generate a binary treatment status indicator (T_i) for each subject from a Bernoulli distribution with subject-specific probability of receiving the experimental treatment equal to $p_{i,treat}$. Those subjects with $T_i = 1$ denote the treated subjects in whom the ATT is defined. We then estimated the propensity score for each subject using a logistic regression model that regresses treatment status (T) on the 6 baseline covariates.

We assumed that the following outcome-selection logistic model related the probability of having the event of interest for subject i ($p_{i,outcome}$) to the six baseline covariates ($X_{1i} - X_{6i}$) and the treatment indicator variable (T_i):

$$\text{logit}(p_{i,outcome}) = \alpha_{0,outcome} + \beta T_i + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_1 X_{4i} + \alpha_2 X_{5i} + \alpha_3 X_{6i}$$

We examined 3 different outcome-selection model scenarios. Under a weak outcome-selection model, the coefficients for the baseline covariates in the above model are as follows: α_1 = log(1.1), $\alpha_2 = \log(1.25)$, and $\alpha_3 = \log(1.5)$. Under a strong outcome selection model, the coefficients for the baseline covariates in the above model are as follows: α_1 = log(1.75), $\alpha_2 = \log(2)$, and $\alpha_3 = \log(2.5)$. Under a mixed outcome-selection model, the coefficients for the baseline covariates in the above model are as follows: α_1 = log(2), $\alpha_2 = \log(1.1)$, and $\alpha_3 = \log(1.5)$. The data-generating process was designed to induce a specific average treatment effect for the treated (ATT), which is the measure of effect that is estimated when propensity-score matching is used. The intercept $\alpha_{0,outcome}$ was chosen so that the true event rate would be approximately 20% if all experimental treated subjects received the control treatment (p_C = 0.2). The coefficient β of the treatment indicator variable was chosen so that the true risk difference due to treatment for experimental treated subjects (i.e. ATT) was $p_C - p_E = -0.1$ under the null hypothesis, and $p_C - p_E = 0$ under the alternative hypothesis, with non-inferiority margin δ_0 = 0.1. The reader is referred to Appendix A.1 of this chapter for a more detailed explanation of how the values of $\alpha_{0,outcome}$ and β were determined. Using the above outcome logistic model, we calculate the subject-specific probability of the outcome ($p_{i,outcome}$). We generate a dichotomous outcome variable (Y_i) for each subject from a Bernoulli distribution with subject-specific probability of having the event of interest equal to $p_{i,outcome}$.

We performed nearest neighbor matching without replacement in order to obtain a propensity-score matched sample. Nearest neighbor matching uses a “greedy” algorithm that makes optimal decisions at each step without attempting to make the best overall (i.e. global) decision. Nearest neighbor matching is one of the most common algorithms used for matching, and it is the easiest to implement and understand (Stuart, Matching Methods for Causal Inference: A Review and a Look Forward, 2010). In nearest neighbor matching without replacement, both experimental and control subjects are first randomly sorted. Then, the first experimental subject is selected to find up to M ($M \geq 1$) of its closest controls in terms of the propensity score (or logit of the scores). This procedure is repeated for all the experimental subjects. Once a control subject has been selected to be matched to a given experimental subject that control subject is no longer available for consideration as a potential match for subsequent experimental subjects. In order to avoid poor matches, one could use nearest neighbor matching with a caliper. It is the same as nearest neighbor matching, but with the further restriction that the absolute difference in the propensity scores (or logit of the propensity scores) of matched subjects must be below some pre-specified threshold (i.e. caliper distance). Thus, when using nearest neighbor matching with a caliper, it is possible that an experimental subject cannot be matched to a control subject. If many experimental treated subjects do not receive a match, then it will be difficult to interpret the estimate of the ATT (Stuart, Matching Methods for Causal Inference: A Review and a Look Forward, 2010). In this simulation study, we examined scenarios that used nearest neighbor matching with a caliper and scenarios that did not use a caliper. For scenarios that used matching with a caliper, we examined 4 different caliper matching scenarios. In the first caliper matching scenario, we match on the logit of the estimated propensity score using a caliper of width equal to

0.2 of the standard deviation of the logit of the propensity score. This caliper width has been shown to result in optimal estimation of risk differences in variety of settings (Austin, Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies, 2011). In the second caliper matching scenario, we match on the logit of the estimated propensity score using a caliper of width equal to 0.5 of the standard deviation of the logit of the propensity score. In the third caliper matching scenario, we match on the raw estimated propensity score using a caliper of width equal to 0.1. In the fourth caliper matching scenario, we match on the raw estimated propensity score using a caliper of width equal to 0.25. We used the *%PSMatching* SAS macro to implement the matching algorithm (Coca-Perraillon, 2007). We also examined scenarios where we set the maximum number of controls matched to each experimental subject to either 1, 2, or 5.

2.3.2. Statistical analyses

For each simulated data set, we performed the following non-inferiority hypothesis test on the propensity-score matched sample:

$$\mathrm{H}_0\colon p_C - p_E \leq \delta_0 \quad \text{vs.} \quad \mathrm{H}_1\colon p_C - p_E > \delta_0$$

where p_C and p_E are the true event probabilities under the control and experimental treatments, respectively, for subjects in whom the ATT is defined, and $\delta_0 < 0$ is the non-inferiority margin. Without loss of generality, we assume that a smaller success probability denotes greater safety. In the propensity-score matched sample, we used two statistical methods to perform the non-inferiority test. First, we used a method proposed by Farrington and Manning (Farrington & Manning, 1990) that assumes independent treatment groups based on the restricted maximum likelihood estimation (MLE):

$$Z_{FM} = \frac{\hat{p}_C - \hat{p}_E - \delta_0}{\sqrt{\widehat{\text{Var}}_{FM}(\hat{p}_C - \hat{p}_E - \delta_0)}} \tag{1}$$

where

$$\widehat{\text{Var}}_{FM}(\hat{p}_C - \hat{p}_E - \delta_0) = \frac{\widetilde{p_C}(1-\widetilde{p_C})}{N_C} + \frac{\widetilde{p_E}(1-\widetilde{p_E})}{N_E}.$$

$\hat{p}_C$ is the proportion of control subjects in the matched sample with the event of interest, $\hat{p}_E$ is the proportion of experimental subjects in the matched sample with the event of interest, N_c and N_E are the number of control and experimental subjects, respectively, in the matched sample.

$$\tilde{p}_C = 2u_{fm}\cos(w_{fm}) - b_{fm}/3a_{fm}, \qquad \tilde{p}_E = \tilde{p}_C - \delta_0$$

where

$$\theta = \frac{\mathrm{N_E}}{N_C}, \qquad a_{fm} = 1 + \theta, \qquad b_{fm} = -\big(1 + \theta + \hat{p}_C + \theta\hat{p}_E + \delta_0(\theta + 2)\big),$$

$$c_{fm} = (-\delta_0)^2 + \delta_0(2\hat{p}_C + \theta + 1) + \hat{p}_C + \theta\hat{p}_E, \quad \mathrm{d_{fm}} = -\delta_0 * (1 + \delta_0)\hat{p}_C,$$

$$u_{fm} = sign(v_{fm})\sqrt{\left(\frac{b_{fm}^2}{(3a_{fm})^2} - \frac{c_{fm}}{3a_{fm}}\right)},$$

$$v_{fm} = \left(\frac{\mathrm{b_{fm}^3}}{(3\mathrm{a_{fm}})^3}\right) - \left(\frac{\mathrm{b_{fm}c_{fm}}}{6\mathrm{a_{fm}^2}}\right) + \left(\frac{\mathrm{d_{fm}}}{2\mathrm{a_{fm}}}\right), \quad \mathrm{w_{fm}} = (\pi + \cos^{-1}\left(\frac{v_{fm}}{u_{fm}^3}\right))/3.$$

$\tilde{p}_C$ and $\tilde{p}_E$ are the restricted MLEs of $\mathrm{p_C}$ and $\mathrm{p_E}$ under the null hypothesis that the risk difference equals δ_0. The test statistic $\mathrm{Z_{FM}}$ is asymptotically standard normal under the null hypothesis.

Second, we used a method proposed in chapter one of this dissertation which adjusts the Farrington-Manning test statistic (1) to account for the potential correlation within matched subjects:

$$Z_{AFM} = \frac{\hat{p}_C^* - \hat{p}_E - \delta_0}{\sqrt{\widehat{\mathrm{Var}}_{AFM}(\hat{p}_C^* - \hat{p}_E - \delta_0)}} \tag{2}$$

where

$$\widehat{\mathrm{Var}}_{\mathrm{AFM}}(\hat{p}_C^* - \hat{p}_E - \delta_0) = \frac{\tilde{p}_C(1-\tilde{p}_C)}{N_C} VIF_C + \frac{\tilde{p}_E(1-\tilde{p}_E)}{N_E} - 2Cov(\hat{p}_C, \hat{p}_E).$$

$\hat{p}_C^*$ is the weighted event probability for the control group in the matched sample, where each control receives a weight that is the inverse of the number of controls matched to the same experimental subject (Stuart, Matching Methods for Causal Inference: A Review and a Look Forward, 2010) (Austin, Statistical Criteria for Selecting the Optimal Number of Untreated Subjects Matched to Each Treated Subject When Using Many-to-One Matching on the Propensity Score, 2010) (Austin, Assessing balance in measured baseline covariates when using many-to-one matching on the propensity-score, 2008). The proposed statistic uses Eldridge et al's (Eldridge, Ashby, & Kerry, 2006) formula to calculate the variance inflation factor (VIF) which is used to adjust the variance of the control group's event rate to account for the correlation among controls in the same matched set. The formula takes into account the possibility of unequal set sizes (i.e. variable-ratio matching, where the number of controls matched to each treated subject is allowed to vary from one matched set to another) by using the coefficient of variation of the set size,

$$VIF_C = 1 + \left(\left[\left\{cv_C^2 \frac{(K-1)}{K} + 1\right\} \overline{n_C}\right] - 1\right) ICC_C \,,$$

where $cv_C = \frac{s_{n_C}}{\overline{n_C}}$ is the coefficient of variation of the set size in the control group. The variance of the set size in the control group is $s_{n_C}^2 = \frac{\sum_{k=1}^{K}(n_{Ck} - \overline{n_C})^2}{(K-1)}$. K is the number of matched sets. $\overline{n_C}$ is the average number of control matched to each treated subject. n_{Ck} is the number of controls in matched set *k*. Wu et al (Wu, Crespi, & Wong, 2012) presented several methods for estimating the intra-class correlation coefficient for binary responses. The proposed statistic uses a modified version of Fleiss-Cuzick estimator for the intra-class correlation coefficient (ICC_C) of the control group. The modified Fleiss-Cuzick estimator is based on Farrington-Manning's restricted MLE of the event rate in the control group instead of the unrestricted MLE. The formula for the ICC_C is

$$ICC_C = 1 - \frac{\sum_{k=1}^{K} \frac{x_{Ck}(n_{Ck} - x_{Ck})}{n_{Ck}}}{(N_C - K)\widetilde{p_C}(1 - \widetilde{p_C})}$$

where x_{Ck} is the number of events in the control group in matched set *k*. To account for the possible correlation between subjects in the control group and experimental group who are matched, the proposed statistic uses a modified version of Obuchowski's method to estimate the covariance between the event rates' of the control group and experimental group. The proposed statistic uses Farrington-Manning's RMLEs ($\tilde{p}_C$ and $\tilde{p}_E$) in estimating the covariance:

$$Cov(\widehat{p_C}, \widehat{p_E}) = \frac{K}{K-1} \sum ((x_{Ck} - n_{Ck}\tilde{p}_C)(x_{Ek} - \tilde{p}_E))/(N_C N_E)$$

A continuity correction may improve the normal approximation of the above proposed statistic for data with very small sample sizes or very highly correlated matched subjects. The proposed test statistic with the continuity correction is

$$Z_{AFM_{CC}} = \frac{\hat{p}_C^* - \hat{p}_E - \delta_0 - CC}{\sqrt{\widehat{\mathrm{Var}}_{AFM}(\hat{p}_C - \hat{p}_E - \delta_0)}} \qquad (3)$$

where

$$CC = \frac{1}{2}\left(\frac{1}{N_{Total}^{eff}}\right), \qquad N_{Total}^{eff} = K * (\rho) + \left(N_E + N_c^{eff}\right)(1-\rho),$$

$$N_C^{eff} = \frac{N_C}{VIF_C}, \qquad \rho = \rho(\widehat{p_C}, \widehat{p_E}) = \frac{Cov(\widehat{p_C}, \widehat{p_E})}{\sqrt{\frac{\widetilde{p_C}(1-\widetilde{p_C})}{N_C} * VIF_C}\sqrt{\frac{\widetilde{p_E}(1-\widetilde{p_E})}{N_E}}}$$

We used the estimated risk difference and standard error of the risk difference based on each of the aforementioned methods to also estimate 95% confidence intervals of the risk difference.

For each of the different scenarios, we simulated 25,000 data sets. When the true risk difference was equal to the non-inferiority margin δ_0 = -0.1 (under null hypothesis), we estimated the empirical type I error rate as the proportion of simulated data sets in which the null hypothesis was rejected with a significance level of less than 0.025. Owing to our use of 25,000 simulated data sets, an empirical type I error rate that was less than 0.023 or greater than 0.027 would be classified as being significantly different from 0.025. When the true risk difference was equal to 0 (under alternative hypothesis), with non-inferiority margin δ_0 = -0.1, we estimated empirical power as the proportion of simulated data sets in which the null hypothesis was rejected with a significance level of less than 0.025. For each of the scenarios, we also estimated the empirical coverage of the 95% confidence intervals of the risk difference as the proportion of estimated 95% confidence intervals that contained the true risk difference. We also determined the mean width of the estimated 95% confidence intervals across the 25,000 simulated data sets. Also, the bias (i.e. $\text{Estimated RD} - \text{True RD}$) of the unweighted (used in the FM method) and weighted (used in the AFM method) estimates of the risk difference were computed across the 25,000 simulated data sets. Bias < 0 indicates bias towards the null.

Finally, we compared the standard deviation of the empirical sampling distribution of the estimated risk difference (i.e. the standard deviation of the 25,000 estimated risk differences across the simulated data sets) with the mean of the estimated standard errors of the estimated risk difference.

2.4. Monte Carlo Simulation Study - Results

Tables 2.1-2.10 report the simulation results for scenarios in which the total sample size (pre-matched data) is 20,000. Tables 2.1-2.6 report the simulation results for the following covariate scenarios using a caliper width equal to 0.2 of the standard deviation (SD) of the logit propensity score: independent normal and Bernoulli covariates, correlated normal and Bernoulli covariates, independent normal covariates, correlated normal covariates, independent Bernoulli covariates, and correlated Bernoulli covariates. Tables 2.7-2.10 report the simulation results for the following caliper scenarios using independent normal and Bernoulli covariates: caliper width of 0.1 based on raw propensity score, caliper width equal to 0.5 of the SD of logit propensity score, caliper width 0.25 based on raw propensity score, and no caliper. Tables 2.11-2.20 report the simulation results for the same scenarios reported in Tables 2.1-2.10, but for a total sample size (before matching) of 2,000.

Each table is split into 3 parts. Part "A" reports the empirical type I error, bias of the estimated risk difference (i.e. ATT), and the ratio of the mean estimated standard error of the risk difference (RD) to the standard deviation of the empirical estimated RD. Part "B" reports the empirical power. Part “C” reports the coverage and width of the empirical 95% confidence intervals of RD under the null and alternative hypotheses. Each table reports the maximum number of controls that were matched to an experimental subject, the mean and standard deviation of the number of controls matched to an experimental subject, and the mean percentage of experimental treated subjects that were successfully matched. Within each table, we report results for 15 scenarios (5 outcome/treatment selection models x 3 scenarios regarding the maximum number of controls matched to each experimental treated subject). We examined scenarios where we set the maximum number of controls matched to each experimental subject to either 1, 2, or 5.

When all or almost all experimental treated subjects were successfully matched to one or more controls with similar propensity scores, the use of a method for correlated matched samples (i.e. the adjusted Farrington-Manning test, AFM) provided empirical type I error rates that were closer to the nominal 2.5% level compared to the use of a method for independent samples (i.e. the original Farrington-Manning test, FM). In all scenarios, the FM test was more conservative than the AFM test when assessing non-inferiority on a propensity-score matched sample. When baseline covariates were strongly related to the outcome, the AFM test provided empirical type I error rates that were satisfactorily close to the nominal 2.5% level, while the FM test resulted in type I error rates that were significantly less than the nominal level.

The empirical type I error rates of the FM test decreased (became very conservative) as the variability (i.e. standard deviation) of the number of controls matched to each experimental subject increased, while the empirical type I error rate of the AFM test stayed close to the nominal level. Unlike the AFM method, the FM method is based on an unweighted estimate of the risk difference, which does not account for one-to-many variable ratio matching. Under variable ratio matching, unweighted estimates of the risk difference were biased toward the non-inferiority null hypothesis. The variability of the number of controls matched to each experimental subject increased when the treatment-selection process was strong (i.e. baseline covariates are strongly related to the treatment assignment), when the covariates were correlated to one another, when the maximum number of controls per experimental treated subject increased, and/or when the sample size in the control group was not sufficiently large.

Empirical type I error rates for the AFM test increased as the mean percentage of experimental treated subjects that were successfully matched to a control subject decreased. The proportion of experimental subjects who were successfully matched decreased when baseline covariates were strongly related to the treatment assignment, when the covariates were correlated to one another, and/or when the sample size in the control group was not sufficiently large. Not including all (or almost all) of the experimental subjects resulted in estimates of the risk difference (i.e. ATT) that are biased towards the non-inferiority alternative hypothesis.

Type I error rates for FM and AFM tests decreased as the caliper increased or when a caliper was not used. Widening the caliper or not using a caliper resulted in estimates of the risk difference that were biased towards the non-inferiority null hypothesis. This effect on type I error rates was more pronounced as the number of controls matched to each experimental subject increased. Matching using a caliper width of 0.1 based on the raw propensity score produced similar results as when matching using a caliper width equal to 0.2 times the standard deviation of the logit of the propensity score.

With respect to power, differences between the FM and AFM tests were greater when covariates were strongly related to the outcome, with the AFM test being more powerful than the FM test. Power for the AFM test increased as the number of controls matched to each experimental subject increased. However, this was not always the case for the FM test. Since the FM test is based on an unweighted estimate of the risk difference which does not account for variable ratio matching, the power of the FM test decreased as the variability of the number of controls matched to each experimental subject increased.

With respect to empirical coverage of the 95% confidence intervals (CI) of RD under the null, differences between the FM and AFM tests were greater when covariates were strongly related to the outcome, with the coverage rates for the CI based on AFM method being more closer to the advertised rate of 95% than for the CI based on the FM method. The relative difference between the widths of the CI was greater when there was a strong outcome-selection process (i.e. covariates strongly related to the outcome) compared with when there was a weak outcome-selection process, with the CI based on the FM method being wider than the CI based on the AFM method.

In general, under the null hypothesis, the estimates of the standard error based on the FM method overestimated the sampling standard deviation of the estimated risk difference when there was a strong outcome-selection process compared with when there was a weak outcome-selection process. In contrast, standard error estimates obtained using the AFM method were closer to the sampling standard deviation of the estimated risk difference, except when there was a relatively large proportion of experimental treated subjects not included in the matched sample.

2.5. Example

Consider a non-randomized cardiovascular device study. In this study, we collected data from a single-arm trial of an experimental embolic protection device for patients with small vessels. We also obtained data from historical control patients who met the inclusion and exclusion criteria of the single-arm trial of the experimental device. The primary objective of this non-randomized study was to assess non-inferiority via risk difference of the experimental device to the active control with respect to major adverse cardiovascular event (MACE). The experimental device is considered non-inferior to active control when the former is no more than 4% higher than the latter with respect to rate of MACE. This leads to the following hypothesis test: $\mathrm{H}_0\colon p_c - p_E \leq -0.04$ vs. $\mathrm{H}_1\colon p_C - p_E > -0.04$. Since subjects in this study were assigned to experimental or active control device in a non-randomized manner, we expected differences between treatment groups with respect to distribution of baseline characteristics. To account for this, the primary analysis was conducted on the propensity-score matched sample using nearest neighbor matching with a caliper width of 0.2 of the standard deviation of the logit of the propensity score. Each experimental subject was matched to 1 or 2 control subjects without replacement. The following are the baseline characteristics used to generate the propensity score: age (years), diameter stenosis (%), diabetes status (yes/no), lesion length (mm), reference vessel diameter (mm), and sex (male/female). The propensity matched sample consisted of 111 subjects treated with the experimental device and 213 treated with the control device. About 1.9 controls were matched to each experimental subject. The unweighted estimate of the risk difference ($\widehat{RD} = \hat{p}_C - \hat{p}_E$) was 0.0442 (or 4.42%). In order to account for the variable ratio matching, we calculated a weighted estimate of the risk difference ($\widehat{RD}^*$), where each control subject's weight is the inverse of the number of control subjects matched to the same experimental subject, while the weight for the experimental subject would be 1. The weighted estimate of the risk difference was $\widehat{RD}^* = 0.0450$ (or 4.50%). When using a method for independent samples, i.e. Farrington-Manning's (FM) method, the estimated standard error of the risk difference was SE_{FM}= 0.034096 and the test statistic value was

$$Z_{FM} = \frac{\widehat{RD} - \delta_0}{SE_{FM}} = \frac{0.0442 - (-0.04)}{0.034096} = 2.47$$

When using a method for correlated matched samples, i.e. the adjusted Farrington-Manning's (AFM) method, the estimated standard error of the risk difference was SE_{AFM}= 0.031332 and the test statistic value was

$$Z_{AFM} = \frac{\widehat{RD}^* - \delta_0}{SE_{AFM}} = \frac{0.045 - (-0.04)}{0.031332} = 2.71$$

where

$$SE_{AFM} = \sqrt{Var(\hat{p}_C) * VIF_C + Var(\hat{p}_E) - 2Cov(\hat{p}_C, \hat{p}_E)}.$$

The unadjusted variance in the control group was $Var(\hat{p}_C)$=0.00029648, variance inflation factor for control group was VIF_C=0.916938, variance in the experimental group was $Var(\hat{p}_E)$=0.00086605, and covariance $Cov(\hat{p}_C, \hat{p}_E)$=0.000078108. The estimated standard error based on the method for correlated matched samples (i.e. AFM) was smaller than the estimated standard error based on the method assuming independent samples (i.e. FM). Both test statistics were greater than the critical value of 1.96. Therefore, based on both test statistics, we reject the null hypothesis and conclude that the experimental device is non-inferior to the active control. The adjusted test statistic (AFM) was larger than the unadjusted test statistic (FM).

2.6. Discussion

We compared a method for independent samples (i.e. Farrington-Manning method) to a method for correlated matched samples (i.e. adjusted Farrington-Manning method) when assessing non-inferiority via risk differences in one-to-many propensity-score matched samples. When all or almost all the experimental treated subjects are well-matched to one or more controls with similar propensity scores, we found that compared to using the Farrington-Manning method, the use of the adjusted Farrington-Manning method resulted in: empirical type I error rates that were closer to the nominal level, higher empirical power, empirical coverage rates of 95% CI that were closer to the advertised 95% rate, and estimated standard errors that were more similar to the sampling standard deviation of the estimated risk difference.

The results from our Monte Carlo simulation study suggests that a method for correlated matched samples is preferable than a method for independent samples when assessing non-inferiority in well-matched samples that contain almost all of the experimental treated subjects. For some scenarios, the original Farrington-Manning test was too conservative (i.e. significantly less than the nominal level), especially when baseline covariates were strongly related to the outcome. Using the adjusted Farrington-Manning method may increase the precision of the estimate of the risk difference and thus increase the empirical power. With respect to empirical type I error rates and power, differences between the FM and AFM methods were greater when the covariates were strongly related to the outcome, when the variability of the number of controls matched to each experimental increased, or when the sample size of the control group was not large enough to find "good" (with respect to propensity score) matches for the experimental subjects.

Using one-to-many matching, instead of one-to-one matching, may increase power and precision, especially when using a method that accounts for correlated matched samples (i.e. AFM method). However, if one is unable to successfully match all (or almost all) of the experimental subjects to at least one control subject, then one should consider another approach (e.g. stratification, regression adjustment, etc.) to reduce confounding. Type I error rates for the AFM tended to be significantly inflated when some of the experimental treated subjects were excluded in the final matched sample. Not including all (or almost all) of the experimental treated subjects in the matched sample resulted in bias estimates of the average treatment effect for the treated (ATT). Widening the caliper may allow more experimental treated subjects to be matched. However, widening the caliper, or not using a caliper at all, may result in experimental subjects to be matched with dissimilar control subjects with respect to propensity scores, resulting in bias estimates of the treatment effect.

ATT is the measure of effect that is estimated when propensity-score matching is used (Imbens, 2004). The true ATT is defined as the average of the within-pair (i.e. 1-to-1) differences of the outcome under the experimental and the potential outcome under the control for each experimental treated subject (Austin, Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies, 2011). Therefore, we recommend using the weighted RD to estimate the ATT when matching one or more controls to each experimental treated subject (i.e. 1-to-≤M matching), where each control subjects' weight is the inverse of the number of control subjects matched to the same experimental subject. The weights are in place to make sure the sample is representative of the population of interest. Using a method for independent samples (i.e. FM method) does not account for the one-to-many variable-ratio matching, which can result in bias estimates of the risk difference.

The current study complements prior published research. Earlier studies found that in many settings, methods for matched samples tended to result in improved inference compared to methods for independent samples when used for analyzing propensity-score matched samples. In prior studies, type I error rates and empirical coverage rates of confidence intervals were studied for differences in means, relative risks, odds ratios, hazard ratios, and/or risk differences (Austin, Type I error rates, coverage of confidence intervals, and variance estimation in propensity-score matched analyses, 2009; Austin, Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples, 2011). All these studies only examined scenarios with one-to-one matching. These prior studies did not examine inferences about non-inferiority in one-to-many propensity-score matching. Also, these prior studies did not study the empirical power of these methods.

In conclusion, if all or almost all the experimental treated subjects are successfully matched to one or more controls with similar propensity scores, then we recommend that the adjusted Farrington-Manning method be used when assessing non-inferiority via risk difference.

3. ASSESSING NON-INFERIORITY VIA RISK DIFFERENCE IN PROPENSITY-SCORE MATCHED STUDIES WHEN THERE ARE MORE EXPERIMENTAL SUBJECTS THAN CONTROL SUBJECTS

3.1. Introduction

Non-inferiority safety studies are often conducted when a new safety concern has arisen for a marketed experimental treatment (Jiang & Xia, 2015). The goal of these studies is to show that the experimental treatment does not result in a clinically unacceptable increase in harm compare to a control treatment. Since dichotomous outcomes are common in clinical research, these non-inferiority safety trials are often designed (i.e. powered) to rule out excess harm based on the risk difference.

Randomized controlled trials (RCTs) are considered the gold standard for assessing non-inferiority of an experimental treatment compared to a control treatment. However, for ethical or practical reasons, it is sometimes not possible to conduct RCTs. Therefore, non-randomized (observational) studies are commonly used to address safety questions for medical drugs, biologics, and devices in a regulatory setting (Levenson & Yue, 2013). However, in non-randomized trials, there is a likelihood of imbalance between treatment groups on clinically relevant baseline characteristics because treatment selection is often influenced by subject characteristics. This could potentially lead to biased estimates of the treatment effect.

Matching experimental subjects to control subjects on baseline covariate information (through a propensity score for example) is a common technique to reduce the bias in treatment effect that is caused by the lack of randomization. Often investigators are confronted with studies where there are a limited number of experimental subjects and a larger (usually much larger) number of control subjects. In these studies, each subject in the experimental group is matched to one or more subjects in the control group. However, in post-market safety observational studies, it is possible that more subjects are treated with the experimental treatment than with the comparator. For example, in a non-randomized post-market safety study to assess the non-inferiority of drug-eluting stents to bare-metal stents on the incidence of MACCE (major adverse cardiovascular and cerebrovascular event) at 30 months, there were more subjects treated with drug-eluting stents (i.e. experimental treatment) than with bare-metal stents (i.e. control treatment) (Kereiakes, Yeh, Massaro, Driscoll-Shempp, & et al., 2015). For these studies, one could match each subject in the control group to one or more subjects in the experimental group.

The most common (and simplest) approach to matching is one-to-one matching, in which pairs of experimental and control subjects are formed. However, there are some drawbacks of using one-to-one matching, especially when there are more experimental treated subjects than controls subjects. First, using one-to-one matching when there are more experimental subjects than control subjects will result in the exclusion of some experimental treated subjects from the final outcome analysis. The exclusion of experimental subjects is usually discouraged when conducting clinical research in a regulatory setting (Levenson & Yue, 2013). Second, if many experimental treated subjects are excluded from the final analysis, then it will be difficult to interpret the estimate of the average treatment effect for the treated (ATT), which is typically the effect measure of interest when matching is used. Third, generally, one-to-one matching is wasteful of sample size and has lower statistical power compared to one-to-many matching. Alternatively, matching multiple experimental subjects to each control subject will generally increase the number of experimental subjects in the final matched sample and, thus, power.

After the matching has created experimental and control groups with adequate balance, we can use the matched sample to assess non-inferiority of the experimental treatment. In chapter one of this dissertation, we presented several methods that could be used to assess non-inferiority for correlated one-to-many matched data, in which many control subjects are matched to each experimental subject. We showed through Monte Carlo simulation that for various correlation patterns the sizes of tests developed for clustered matched pair data and GEE-based tests are inflated when applied to the case where many control subjects are matched to each experimental subject. On the other hand, the size of the test that we proposed in chapter one (i.e. adjusted Farrington-Manning test) stays close to the nominal level for a variety of correlation patterns when many control subjects are matched to each experimental subject.

Although it is possible to apply the aforementioned methods to the case where one or more experimental subjects are matched to each control, it is unclear how well these methods will perform with respect to type I error and power in this setting. Through a simulation study of independent (non-matched) data, Dann and Koch (2008) found that the empirical type I error of the Farrington-Manning test statistic became progressively higher so as to exceed the nominal level as relatively more sample size is placed in the experimental group than in the control group (experimental >> controls). Dann and Koch (2008) found that the appropriate method to use for assessing non-inferiority depends on the sample size allocation of subjects in the experimental and control groups. This statistical finding by Dann and Koch (2008) raises the question of whether this phenomenon also applies to the proposed adjusted Farrington-Manning test statistic. It is unclear how sample size allocation (experimental >> controls vs. experimental << controls) will impact the operating characteristics of statistical methods that are used to assess non-inferiority for one-to-many correlated matched data.

The objective of the current study is to assess the performance of non-inferiority tests for correlated one-to-many matched data, in which many experimental subjects are matched to each control. We also want to compare the effect on assessing non-inferiority when a method for independent samples is used compared with a method for correlated matched samples is used in a one-to-many propensity-score matched sample, wherein many experimental subjects are matched to each control. Section 2 provides notation and a review of non-inferiority tests for correlated matched data. Section 3 gives an evaluation of the performance of each of the non-inferiority tests in terms of the empirical type I error and power by simulations under various general correlation scenarios and propensity-score scenarios. Section 4 provides an example of assessing non-inferiority of a cardiovascular device in a non-randomized post-market surveillance study using one-to-many propensity-score matching, in which there are more experimental subjects than control subjects. Section 5 contains a discussion and conclusion.

3.2. Methods

In this section, we provide a review of non-inferiority tests for correlated matched data and how these tests can be applied to the case where one or more experimental subjects are matched to each control subject. The reader is referred to chapter one of this dissertation for a more detailed background and explanation of these methods.

In chapter one of this dissertation, we argue that it is possible to apply non-inferiority tests for clustered matched pair data to one-to-many matched data by treating each 1: M matched set as a cluster of 1:1 matched data. For example, a matched set with a control subject matched to 3 experimental subjects could be treated as a *cluster* of *three* 1-to-1 matched data by simply pairing the binary response of the control subject to the binary response of *each* matched experimental subject. When our data is treated as matched-pair data, there are four possible pairs of responses, i.e. (1,1), (1,0), (0,1), and (0,0), where the first element is the response to the control procedure and the second element is the response to the experimental procedure. Let $b_k \; and \; c_k$ be the observed frequencies of (1,0) and (0,1), respectively, in cluster *k*. Clearly, b_k and c_k are frequencies of discordant pairs. By treating each matched set as a cluster of one-to-one matched data, we argue that it is possible to use non-inferiority tests for clustered matched pair data to assess non-inferiority when many experimental subjects are matched to each control.

Durkalski et al (2003) proposed the following non-inferiority test statistic for clustered matched pair data based on method of moments:

$$Z_{MOM} = \sum_{k=1}^{K}\left(\frac{b_k - c_k}{n_k} - \delta_0\right) / \left[\sum_{k=1}^{K}\left(\frac{b_k - c_k}{n_k} - \delta_0\right)^2\right]^{\frac{1}{2}}. \tag{1}$$

Let K be the total number of matched sets. Let n_k be the number of subjects from the experimental group that are matched to subject k from the control group, where k=1, …, K. Let $\delta_0 < 0$ be the non-inferiority margin.

Nam and Kwon (2009) proposed an ICC (intraclass correlation coefficient) adjusted score statistic to test non-inferiority for clustered matched-pair binary data:

$$Z_{ICC} = [\sum_{k=1}^{K}(b_k - c_k) - N_E\delta_0]/[N_E(\tilde{p}_{10} + \tilde{p}_{01} - \delta_0^2)\bar{c}]^{1/2} \;, \tag{2}$$

where N_E is the total number of experimental subjects in the matched sample, $\tilde{p}_{10}$ and $\tilde{p}_{01}$are the restricted maximum likelihood estimators of p_{10}(the response probabilities for the discordant pair (1,0)) and p_{01}(the response probabilities for the discordant pair (0,1)), and $\bar{c}$ is the variance inflation factor.

Based on sampling techniques, Obuchowski (Obuchowski, 1998) proposed a test statistic which takes into account correlation within a cluster. Nam and Kwon (Nam & Kwon, Non-inferiority tests for clustered matched pair data, 2009) extended Obuchowski's statistic to test for non-inferiority in clustered matched-pair data and proposed

$$\mathrm{Z_O} = (\hat{p}_C - \hat{p}_E - \delta_0)/\left[\widehat{\mathrm{Var}}(\hat{p}_C - \hat{p}_E - \delta_0)\right]^{\frac{1}{2}}, \tag{3}$$

where $\hat{p}_C$ and $\hat{p}_E$ are the event rates in the control group and experimental group, respectively, and $\widehat{\mathrm{Var}}(\hat{p}_C - \hat{p}_E - \delta_0) = \mathrm{Var}(\hat{p}_C) + \mathrm{Var}(\hat{p}_E) - 2\mathrm{Cov}(\hat{p}_C, \hat{p}_E)$. The variance is estimated using sampling techniques introduced by Obuchowski.

Instead of applying the aforementioned tests to correlated one-to-many matched data, one could potentially use generalized estimating equations (GEE) for comparing correlated proportions. In order to estimate the treatment effect, one could run a GEE model with the dichotomous outcome as the dependent variable and treatment group as the independent variable. One could use an identity link function and assume a binomial distribution for the outcome in the GEE model. The coefficient of the treatment group variable in the GEE model will be the estimated risk difference β. To account for the possible correlation among matched subjects, one could specify an independent, exchangeable, or unstructured correlation structure in the GEE model. Using the estimated risk difference and its standard error, we are able to derive a test statistic to assess non-inferiority, given a pre-specified working correlation structure *s*, where *s*={independence, exchangeable, or unstructured} (Mascha & Sessler, 2011):

$$Z_s^{GEE} = (\beta - \delta_0)/SE(\beta) \quad (4)$$

Farrington and Manning (Farrington & Manning, 1990) proposed the following non-inferiority test statistic assuming independent treatment groups based on the restricted maximum likelihood estimation

$$Z_{FM} = \frac{\hat{p}_C - \hat{p}_E - \delta_0}{\sqrt{\widehat{\mathrm{Var}}(\hat{p}_C - \hat{p}_E - \delta_0)_{FM}}} \quad (5)$$

In order to account for the possible correlation among matched subjects, we proposed a test statistic in chapter one of this dissertation that adjusts the Farrington-Manning test statistic and is appropriate for assessing non-inferiority when each treated subject is matched to one or more control subjects. We extend the proposed test statistic to the setting where each control subject is matched to one or more experimental groups (number of experimental subjects >> number of control subjects):

$$Z_{AFM} = \frac{\hat{p}_C - \hat{p}_E - \delta_0}{\sqrt{\widehat{\mathrm{Var}}_{AFM}(\hat{p}_C - \hat{p}_E - \delta_0)}} \tag{6}$$

where

$$\widehat{\mathrm{Var}}_{\mathrm{AFM}}(\hat{p}_C - \hat{p}_E - \delta_0) = \frac{\tilde{p}_C(1-\tilde{p}_C)}{N_C} VIF_C + \frac{\tilde{p}_E(1-\tilde{p}_E)}{N_E} VIF_E - 2Cov(\hat{p}_C, \hat{p}_E)$$

Let $\tilde{p}_C$ and $\tilde{p}_E$ are the Farrington-Manning's restricted maximum likelihood estimators of the event rates for the control and experimental treatments, respectively, under the null hypothesis that the risk difference equals δ_0. VIF_C and VIF_E are the variance inflation factors for the control group and experimental group, respectively, which take into account the possible correlation among matched subjects in each treatment group. $Cov(\hat{p}_C, \hat{p}_E)$ is Obuchowski's estimate of the covariance between the control and experimental group, which accounts for the possible correlation between matched experimental and control subjects.

For situations in which the sample size is very small or when subjects within a matched set are extremely highly correlated, a continuity correction may be necessary to improve the normal approximation of the proposed statistic. The same continuity correction (CC) mentioned in chapter one can also be applied:

$$Z_{AFM_{CC}} = \frac{\hat{p}_C - \hat{p}_E - \delta_0 - CC}{\sqrt{\widehat{\mathrm{Var}}_{AFM}(\hat{p}_C - \hat{p}_E - \delta_0)}} \tag{7}$$

The reader is referred to chapter one of this dissertation for a more detailed explanation of how to calculate the continuity correction.

3.3. Simulation Study

In this section, we describe two Monte Carlo simulation studies that were conducted to examine the effect on the operating characteristics of non-inferiority tests for correlated matched data when relatively more sample size is placed in the experimental group than in the control group. First, we performed a Monte Carlo simulation study to assess the empirical type I error rate and power of statistics, $Z_{MOM}, Z_{ICC}, Z_O, Z_S^{GEE}, Z_{FM}, Z_{AFM},\ and\ Z_{AFM_{CC}}$ for assessing non-inferiority of an experimental treatment compared with a control treatment for correlated one-to-many matched data, where each control subject is matched to one or more experimental subjects. Second, we compare a method that assumes independent samples to a method that assumes correlated matched data when assessing non-inferiority via risk differences in *propensity-score* matched studies, in which each control subject is matched to one or more experimental subjects. In these simulation studies, we want to test the following hypotheses: $\mathrm{H}_0: p_C - p_E \leq\ \delta_0$ vs $\mathrm{H}_1: p_C - p_E > \delta_0$, where p_C and p_E are the true event probabilities under the control and experimental treatments, respectively, and $\delta_0 < 0$ is the non-inferiority margin.

3.3.1. Data-generating process

For the first simulation study, we used a data-generating process identical to the one used in chapter one of this dissertation and similar to the one used by Nam and Kwon (Nam & Kwon, Non-inferiority tests for clustered matched pair data, 2009). Pre-specified parameters in the simulation study are the number of clusters or matched sets (*K*), the number of experimental subjects matched to a control subject in the *k*th matched set (n_k), the distribution or variability of n_k, the probabilities of having an event for the control and experimental procedures (p_C and p_E, respectively) and the non-inferiority margin (δ_0). We specify two within-matched set correlations: the correlation of the responses in the experimental group (r_E) and the correlation of the responses between the experimental subjects and the control subjects (r_{CE}). The simulation study was conducted using SAS version 9.4 via PROC IML and macro-language. SAS PROC GENMOD was used to run GEE models. For each matched set, we generate a random vector from a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance $\mathbf{\Sigma}$ where correlation coefficients (r_C and r_{EC}) are specified. Denote matched set normal variates as *z*. Then, we generate a binomial response for each procedure. Define *y* as *y*=1 if z ≤ c or *y* = 0 otherwise, where *c* is a cut-off point which satisfies Pr(z ≤ c) = p_C if z is a normal variate from the control group and Pr(z ≤ c) = p_E if z is a normal variate from the experimental group. When the size of the matched sets (n_k) varies, we consider two kinds of distributions: the value of n_k is generated from a uniform distribution or from a beta(2,3) distribution. In our simulation study, 25,000 data sets were generated for each configuration to calculate empirical Type I error rates and power. If the expected Type I error rate of a test is indeed a nominal 0.025 level, then a 95% confidence interval for α=0.025 is (0.023, 0.027) from 25,000 simulated data. Those empirical Type I error rates above the interval are shown with a caret sign (^) to indicate inflated Type I error rates and those below the interval are shown with an asterisk sign (*) to indicate a conservative test. With regards to GEE-based tests, we only report the results from GEE-based tests that assume an exchangeable correlation structure since GEE-based tests that use unstructured correlation structures often produce unstable results (e.g. fail to converge), and GEE-based tests that use independence correlation structures produce similar results to those that use exchangeable correlation structures.

For the second simulation study, we use a data-generating process similar to the one used in chapter two of this dissertation and in a study by Austin (Austin, Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples, 2011). First, we randomly generated a data set for a specified number of subjects. We examined three different scenarios regarding the sample size allocation of subjects treated with the experimental treatment and control treatment. In the first scenario, we generated 2500 subjects, in which approximately 60% (1500 subjects) are exposed to the experimental treatment and 40% (1000 subjects) are exposed to the control treatment, thus resulting in a sample size allocation ratio of 1:1.5. In the second scenario, we generated 3000 subjects, in which approximately 66.7% (2000 subjects) are exposed to the experimental treatment and 33.3% (1000 subjects) are exposed to the control treatment, thus resulting in a sample size allocation ratio of 1:2. In the third scenario, we generated 5000 subjects, in which approximately 80% (4000 subjects) are exposed to the experimental treatment and 20% (1000 subjects) are exposed to the control treatment, thus resulting in a sample size allocation ratio of 1:4. For each subject, we randomly generated three independent standard normal random variables ($X_1 - X_3$) and three Bernoulli random variables ($X_4 - X_6$). The parameter value for all Bernoulli random variables was 0.5.

After generating the six baseline covariates for each subject i ($X_{1i} - X_{6i}$), we then calculate the subject-specific probability of receiving the experimental treatment ($p_{i,treat}$) using the following treatment-selection logistic model under a specified scenario:

$$\text{logit}(p_{i,treat}) = \alpha_{0,treat} + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_1 X_{4i} + \alpha_2 X_{5i} + \alpha_3 X_{6i}.$$

We examined 3 different treatment-selection logistic model scenarios. Under a weak treatment-selection model, the coefficients for the baseline covariates in the above model were as follows: α_1 = log(1.1), $\alpha_2 = \log(1.25)$, and $\alpha_3 = \log(1.5)$. Under a strong treatment-selection model, the coefficients for the baseline covariates in the above model were as follows: α_1 = log(1.75), $\alpha_2 = \log(2)$, and $\alpha_3 = \log(2.5)$. Under a mixed treatment-selection model, the coefficients for the baseline covariates in the above model were as follows: α_1 = log(1.1), $\alpha_2 = \log(2)$, and $\alpha_3 = \log(1.5)$. The intercept $\alpha_{0,treat}$ was chosen so that the desired amount of subjects would be exposed to the experimental treatment (e.g. 60% of the 2500 simulated subjects would be treated with the experimental treatment). The reader is referred to Appendix A.1 in chapter two for a more detailed explanation of how the value of $\alpha_{0,treat}$ was determined. We then generate a binary treatment status indicator (T_i) for each subject from a Bernoulli distribution with subject-specific probability of receiving the experimental treatment equal to $p_{i,treat}$. Those subjects with T_i = 1 denote the treated subjects in whom the average treatment effect for treated (ATT) is defined. We then estimated the propensity score for each subject using a logistic regression model that regresses treatment status (T) on the six baseline covariates.

We assumed that the following outcome-selection logistic model related the probability of having the event of interest for subject i ($p_{i,outcome}$) to the six baseline covariates ($X_{1i} - X_{6i}$) and the treatment indicator variable (T_i):

$$\text{logit}(p_{i,outcome}) = \alpha_{0,outcome} + \beta T_i + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_1 X_{4i} + \alpha_2 X_{5i} + \alpha_3 X_{6i}$$

We examined 3 different outcome-selection model scenarios. Under a weak outcome-selection model, the coefficients for the baseline covariates in the above model are as follows: α_1 = log(1.1), $\alpha_2 = \log(1.25)$, and $\alpha_3 = \log(1.5)$. Under a strong outcome selection model, the coefficients for the baseline covariates in the above model are as follows: α_1 = log(1.75), $\alpha_2 = \log(2)$, and $\alpha_3 = \log(2.5)$. Under a mixed outcome-selection model, the coefficients for the baseline covariates in the above model are as follows: α_1 = log(2), $\alpha_2 = \log(1.1)$, and $\alpha_3 = \log(1.5)$. The data-generating process was designed to induce a specific average treatment effect for the treated (ATT). The intercept $\alpha_{0,outcome}$ was chosen so that the true event rate would be approximately 20% if all experimental treated subjects received the control treatment (p_C = 0.2). The coefficient β of the treatment indicator variable was chosen so that the true risk difference due to treatment for experimental subjects (i.e. ATT) was $p_C - p_E = -0.05$ under the null hypothesis and $p_C - p_E = 0$ under the alternative hypothesis. The pre-specified non-inferiority margin was δ_0 = -0.05. The reader is referred to Appendix A.1 in chapter two of this dissertation for a more detailed explanation of how the values of $\alpha_{0,outcome}$ and β were determined. Using the above outcome logistic model, we calculate the subject-specific probability of the outcome ($p_{i,outcome}$). We generate a dichotomous outcome variable (Y_i) for each subject from a Bernoulli distribution with subject-specific probability of having the event of interest equal to $p_{i,outcome}$.

We examined two matching algorithm scenarios. First, we performed nearest neighbor matching without replacement in order to obtain a propensity-score matched sample. A control subject was matched to one or more experimental treated subjects. We matched on the logit of the estimated propensity score using a caliper of width equal to 0.2 of the standard deviation of the logit of the propensity score. This caliper width has been shown to result in optimal estimation of risk differences in variety of settings (Austin, Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies, 2011). We set the maximum number of experimental subjects matched to each control subject to either 1, 2, or 5. The second matching algorithm we examined was nearest neighbor matching with replacement, where each experimental treated subject was matched to a control with replacement (where a control can be used as a match for more than one treated individual) on the logit of the estimated propensity score using a caliper of width equal to 0.2 of the standard deviation of the logit of the propensity score. We used the *%PSMatching* SAS macro to implement the matching algorithms (Coca-Perraillon, 2007).

In the propensity-score matched sample, we used two statistical methods to perform the non-inferiority test. First, we used a method that assumes independent data (i.e. original Farrington-Manning test statistic). Second, we used a method that accounts for the potential correlation within matched subjects (i.e. Nam-Kwon test statistic for clustered matched pair data applied to one-to-many matched data). In our simulation study, 25,000 data sets were generated for each scenario to calculate empirical Type I error rates, power, and bias.

3.3.2. Simulation Results

Tables 3.1-3.7 reports the results from the first simulation study. Tables 3.1-3.7 summarize the simulated Type I error rates and power for statistics, $Z_{MOM}, Z_{ICC}, Z_O, Z_S^{GEE}, Z_{FM}, Z_{AFM},\ and\ Z_{AFM_{CC}}$ under various combinations of intra (r_E) and inter (r_{CE}) correlations assuming there are 200 matched sets, the event probability under the control treatment is p_C = 0.2, and non-inferiority margin is δ_0 = -0.1. We investigated scenarios where up to 1, 2, or 5 experimental subjects are matched to each control subject, assuming that the number of experimental subjects matched to each control subject follows either a constant, uniform, or non-uniform (beta) distribution.

From Tables 3.1-3.7, we see that the empirical type I error rate for Farrington-Manning (FM) statistic is not stable across different correlation structures. The FM statistic becomes more conservative as the correlation between experimental subjects and control subjects (r_{CE}) increases. However, if the correlation of the responses between the experimental and control groups is relatively low (r_{CE} < 0.2), empirical Type I error rates for *FM* increases and is above the nominal level as the correlation of the responses in the experimental group (r_E) increases. Also, as can be seen in Table 3.5, the FM statistic is significantly inflated when every control subject is matched to 5 experimental subjects and, at the same time, the data is independent (r_{CE}=0 and r_E=0). This result is consistent with Dann and Koch's (Dann & Koch, 2008) simulation study which found that the FM test produce slightly higher than nominal type I errors when more sample size is placed in the experimental group, when in the context of independent data.

From Tables 3.1-3.7, we see that the GEE_{Exch}, MOM, and O statistics are conservative except when high correlation (r_E=0.8 and/or r_{CE}=0.8) is present. The AFM statistic tends to produce type I errors close to the nominal level except when high correlation is present or when the sample size of the experimental group increases (i.e. the number of experimental subjects matched to each control subject increases). The continuity corrected AFM (AFM_{CC}) statistic does a better job than the AFM in controlling type I error rates at the nominal level when high correlation is present. However, the empirical Type I error rates for AFM_{CC} still increases and, at times is above the nominal level, as more sample size is placed in the experimental group than in the control group. On the other hand, the ICC-adjusted test statistic by Nam and Kwon (Nam & Kwon, Non-inferiority tests for clustered matched pair data, 2009) was conservative (i.e. stayed below the nominal level) for the various correlation structures that were simulated.

From Tables 3.1-3.7, empirical power of each of the test statistics that account for correlated matched data ($Z_{MOM}, Z_{ICC}, Z_O, Z_S^{GEE}, Z_{AFM},\ and\ Z_{AFM_{CC}}$) are fairly similar across different correlation structures. The empirical power of the FM test seems to differ the most with respect to the empirical power of the other statistics when the correlation among matched subjects increases.

Tables 3.8-3.10 report the results from the second simulation study, which examined the effect on statistical inference when using a *propensity-score* matched sample. We compare a method that assumes independent samples (i.e. FM statistic) to a method that assumes correlated matched data (i.e. ICC-adjusted score test by Nam-Kwon) when assessing non-inferiority via risk differences in *propensity-score* matched studies, in which each control subject is matched to one or more experimental subjects. We chose Nam-Kwon's ICC-adjusted test statistic over the other aforementioned statistics because, based on the first simulation study in this chapter, it was found to be conservative for a variety of correlation structures. The other aforementioned test statistics for correlated matched data were anti-conservative for some correlation structures.

Tables 3.8-3.10 summarize the simulated Type I error rates and power for the FM and ICC statistics under various pre-matched sample size allocations for control versus experimental groups including 1:1.5 (Tables 3.8A, 3.8B), 1:2 (Tables 3.9A, 3.9B), and 1:5 (Tables 3.10A, 3.10B). Tables 3.8A, 3.9A, and 3.10A report the empirical type I error, bias of the estimated risk difference (i.e. ATT), and the ratio of the mean estimated standard error of the risk difference (RD) to the standard deviation of the empirical estimated RD. Tables 3.8B, 3.9B, and 3.10B are similar to 3.8A, 3.9A, and 3.10A, but report the empirical power instead of the empirical type I error rates. Tables 3.8-3.10 all report the maximum number of experimental subjects that were matched to a control subject, the mean and standard deviation of the number of experimental subjects matched to a control subject, and the mean percentage of experimental subjects that were successfully matched and included in the final matched sample. Within each table, we report results for 15 scenarios (5 combinations of outcome and treatment selection models x 3 scenarios regarding the maximum number of controls matched to each experimental treated subject). We examined scenarios where we set the maximum number of controls matched to each experimental subject to either 1, 2, or 5.

Tables 3.8-3.10 indicate that the empirical type I error rates of the FM test decreased as the variability (i.e. standard deviation) of the number of experimental subjects matched to each control subject increased. On the other hand, when most of the treated subjects were successfully matched, the empirical type I error rates of the ICC test were closer to the nominal level compared to the FM test even when the variability of the number of experimental subjects matched to each control increased. Unlike the ICC method, the FM method is based on an unweighted estimate of the risk difference, which does not account for variable ratio matching. Under variable ratio matching, unweighted estimates of the risk difference were biased toward the non-inferiority null hypothesis. Nam-Kwon's ICC-adjusted test is based on the weighted risk difference, in which each control subject's outcome is weighted by the number of experimental subjects matched to the same control subject, while the weight for each experimental subject is 1. Table 3.8-3.10 show that if most of the treated subjects were successfully matched, estimates of the risk difference for the treated (i.e. average treatment effect for the treated) were less bias when using weighted risk differences than when using unweighted risk differences.

According to Tables 3.8-3.10, empirical type I error rates for the ICC test increased as the mean percentage of experimental treated subjects that were successfully matched to a control subject decreased. The percentage of experimental subjects that were successfully matched to a control subject was relatively low when baseline covariates were strongly related to the treatment assignment. When the number of treated subjects matched to each control subject was nearly constant (e.g. as in the case of 1-to-1 matching), empirical type I error rates for the FM test also increased as the percentage of experimental subjects that were successfully matched to a control subject decreased. Not including all (or almost all) of the experimental treated subjects resulted in estimates of the risk difference for the treated (i.e. ATT) that were biased towards the non-inferiority alternative hypothesis.

Tables 3.8-3.10 show that the percentage of experimental subjects that were successfully matched to a control subject decreased as relatively more sample size (pre-matching) was in the experimental group compared to the control group. It is difficult to successfully match all the experimental treated subjects if there is a lot more experimental subjects than control subjects. Using one-to-many matching, compared to one-to-one matching, resulted in a higher percentage of treated subjects to be successfully matched.

Tables 3.8-3.10 indicate that in the case where there are more experimental subjects than control subjects, if we decrease the number of experimental subjects that can be matched to each control, then the mean percent of experimental subjects successfully matched also decreases. Having a relatively low percentage of experimental subjects successfully matched to a control will result in biased samples that favor the alternative hypothesis, and thus may inflate type I error rates and power. Differences between the FM and ICC tests were greater with respect to power when covariates were strongly related to the outcome, with the ICC test being more powerful than the FM test.

Table 3.11A provides empirical type I error rates, bias, and estimation of SE of RD results when matching each treated subject to a control with replacement when there are more experimental subjects than controls subjects. Table 3.11B provides the empirical power for the aforementioned matching scenario. We performed the propensity score matching under the following pre-match sample size allocation ratios (treated/control) scenarios: 1.5, 2, 4. Regardless of the pre-match sample size allocation ratio, most of the treated subjects were successfully matched when matching each treated subject to a control with replacement. The percentage of treated subjects that were successfully matched was lowest (97.5%) when treatment-selection model was strong. For various scenarios under this matching algorithm (i.e. matching with replacement), the bias of the estimated risk difference is closer to 0 and Type I error rates are closer to the nominal level when using a method that accounts for the variable ratio matching than when using to a method that does not account for matching.

Matching each treated subject to a control with replacement may result in a higher percentage of treated subjects successfully matched compared to matching each control to one or more treated subjects without replacement. For example, when the pre-match sample size allocation ratio (treated/control) is 1.5, and outcome and treatment selection models are strong, 97.8% (Table 3.11B) of treated subjects were matched when matching each treated subject to a control with replacement vs. 83.5% (Table 3.8B) of treated subjects were matched when matching each control subject to up to 5 controls without replacement. However, matching with replacement may result in the estimate of the treatment effect to be based on a small number of controls (e.g. 43 treated subjects were matched to the same control) resulting in a decrease in power. For example, when the pre-match sample size allocation ratio (treated/control) is 1.5, and outcome and treatment selection models are strong, empirical power for the Nam-Kwon ICC test was 29.2% (Table 3.11B) when matching each treated subject to a control with replacement vs. 62.2% (Table 3.8B) when matching each control subject to up to 5 controls without replacement.

3.4. Example

Consider a non-randomized post-market surveillance study for a cardiovascular stent device. This study was a secondary analysis for the DAPT study (Dual Antiplatelet Therapy) (Mauri, Kereiakes, Yeh, & al., 2014). The primary objective of this secondary analysis was to assess non-inferiority via risk difference of drug-eluting stents (DES) to bare-metal stents (BMS) on the incidence of major adverse cardiovascular and cerebrovascular events (MACCE) at 30 months. The DES was considered non-inferior to BMS if the former was no more than 2.28% higher than the latter with respect to the rate of MACCE. This leads to the following non-inferiority hypothesis test: $\mathrm{H}_0\colon p_{BMS} - p_{DES} \leq -0.0228$ vs. $\mathrm{H}_1\colon p_{BMS} - p_{DES} > -0.0228$. In the DAPT study, the choice of stent type was at the discretion of the physician, and thus, subjects were assigned to DES or BMS stents in a non-randomized manner. There were approximately six times as many DES (i.e. experimental) subjects as BMS (i.e. control) subjects. Since subjects in this study were assigned to DES or BMS stents in a non-randomized manner, we expected differences between treatment groups with respect to distribution of baseline characteristics. To account for this, the analysis was conducted on a propensity-score matched sample using nearest neighbor matching with a caliper width of 0.1. A BMS-treated subject was matched on propensity score to a variable number of DES-treated subjects without replacement, up to a maximum of 8 (a larger number of DES subjects did not appreciably increase power based on the FM test). The propensity score was defined as the probability of receiving the DES (versus BMS) conditioned on the individual's clinically relevant baseline variables. The propensity score was estimated using a logistic regression with treatment as the outcome and 55 independent variables from baseline.

Of the 13,257 DES and 2,056 BMS subjects available for matching, only 8,315 DES and 1,730 BMS subjects were successfully matched when matching up to 8 DES subjects to each BMS subject without replacement and using a caliper width of 0.1. To assess the adequacy of the match, weighted standardized differences in clinical characteristics between groups were calculated (Austin, Assessing balance in measured baseline covariates when using many-to-one matching on the propensity-score, 2008). Treatment groups were balanced with respect to baseline characteristics since standardized differences were <10% for all match variables. The unweighted estimate of the risk difference ($\widehat{RD} = \hat{p}_{BMS} - \hat{p}_{DES}$) was 0.0134 or 1.34% ($\hat{p}_{BMS}$ =12.72% and $\hat{p}_{DES}$ =11.38%). When using a method for independent samples, i.e. Farrington-Manning's (FM) method, the estimated standard error of the risk difference was SE_{FM}= 0.007999 and the test statistic value was

$$Z_{FM} = \frac{\widehat{RD} - \delta_0}{SE_{FM}} = \frac{0.0134 - (-0.0228)}{0.007999} = 4.53$$

In order to account for the variable ratio matching and obtain a less bias estimate of the ATT, a weighted risk difference was calculated. Each BMS subject's weight was the number of DES subjects matched to the same BMS subject, while the weight for each DES subject was 1. The weighted estimate of the risk difference was $\widehat{RD}^{*}$= 0.0267 or 2.67% ($\hat{p}^{*}_{BMS}$ =14.05% and $\hat{p}_{DES}$ =11.38%). In order to account for correlated matched data, we propose treating each one-to-many matched set in our matched sample as a cluster of matched pair data, and then apply Nam-Kwon's ICC-adjusted score test statistic to assess non-inferiority. If we treat each one-to-many matched set as a cluster of matched pair data, then we can say that there are N_{BMS}=K=1,730 clusters (or matched sets) and N_{DES}=8,315 matched pairs in our propensity-score matched sample. Using Nam-Kwon's ICC-adjusted score method, the estimated standard error of the difference between the observed frequencies of the discordant pairs ($b_k - c_k$) was SE_{ICC}= 89.37354 and the test statistic value was

$$Z_{ICC} = \frac{\sum_{k=1}^{K}(b_k - c_k) - N_{DES}\delta_0}{[N_{DES}(\tilde{p}_{10} + \tilde{p}_{01} - \delta_0^2)\bar{c}]^{\frac{1}{2}}} = \frac{N_{DES}\widehat{RD}^{*} - N_{DES}\delta_0}{SE_{ICC}}$$

$$= \frac{8315(0.0267) - 8315(-0.0228)}{89.37354} = 4.61$$

Both test statistics were greater than the Z critical value of 1.96 at the 0.025 significance level. Therefore, based on both test statistics, we reject the null hypothesis and conclude that DES stents is non-inferior to BMS stents.

11870 DES and 1527 BMS subjects were successfully matched when matching each DES subject to a BMS subject with replacement using a caliper width of 0.015 times the standard deviation of the logit of the propensity score. Using the aforementioned caliper width, treatment groups were balanced with respect to baseline characteristics since standardized differences were <10% for all match variables. The weighted estimate of the risk difference was $\widehat{RD}^{*}$= 0.057 or 5.7% ($\hat{p}_{BMS}^{*}$ =17.5% and $\hat{p}_{DES}$ =11.8%). We treat each control subject and its one or more matched experimental treated subjects as a matched set or cluster. Using Nam-Kwon's ICC-adjusted score method, the test statistic value was $Z_{ICC} = 3.90$. Therefore, we reject the null hypothesis and conclude that DES stents is non-inferior to BMS stents at the 0.025 significance level.

3.5. Discussion

Through a simulation study of independent (non-matched) data, Dann and Koch (2008) found that the appropriate method to use for assessing non-inferiority depends on the sample size allocation of subjects in the experimental and control groups. Similarly, the results from our study suggests that the appropriate method to use when assessing non-inferiority for correlated matched data also depends on the sample size allocation. In general, the adjusted Farrington-Manning (AFM) method produced nominal type I error rates when the sample size between the groups were equal or when there were more controls than experimental subjects in the matched sample. However, the type I error rates for the AFM became progressively higher so as to exceed the nominal level as relatively more sample size is placed in the experimental group. The opposite phenomenon occurred when applying the Nam-Kwon ICC-adjusted score test (ICC) for clustered matched pair data to analyze one-to-many matched data. It produced nominal type I error rates when there were more experimental subjects than control subjects, but type I error rates were inflated when more sample size was placed in the control group.

Thus, like with independent data, the choice of an appropriate method for assessing non-inferiority via risk difference for correlated matched data is dependent on the sample size allocation.

If all or almost all the experimental treated subjects are successfully matched to one or more controls with similar propensity scores, then the adjusted Farrington-Manning method (in the case where the number of control subjects is greater than the number of experimental subjects) or the Nam-Kwon method (in the case where the number of experimental subjects is greater than the number of control subjects) can be used to increase power or control the type I error rate at the nominal level when assessing non-inferiority via risk difference. When there are more experimental subjects than control subjects, matching each treated subject to a control with replacement may result in a larger number of treated subjects to be successfully matched compared to matching each control subject to one or more treated subjects without replacement, however maybe at the expense of statistical power since it is possible that the treatment effect estimate be based on just a small number of controls. If one is unable to obtain a well-matched sample that contains almost all of the experimental treated subjects, then one could use other methods to estimate and make inferences about the treatment effect, including stratification on the propensity score, inverse probability weighting using the propensity score, or covariate adjustment using the propensity score.

APPENDIX

A.1. Chapter 1 Tables: Non-Inferiority Tests for One-To-Many Matched Data

Table 1.1. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k=1, n_k~constant, p_C=0.1, δ_0=-0.05

Correlation Structure	*Type I error rate (%)*									*Power (%)*								
r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	2.4	2.7^	2.4	2.5	2.7^	2.7^	2.7^	2.4	2.2*	36.7	38.1	36.7	37.8	39.1	39.1	39.1	36.5	34.8
0.2	2.0*	2.6	2.3*	2.5	2.7^	2.7^	2.7^	2.4	2.2*	36.6	41.2	39.9	41.0	42.6	42.6	42.6	39.9	38.0
0.4	1.5*	2.6	2.4	2.6	2.8^	2.8^	2.8^	2.5	2.2*	36.1	45.2	43.9	45.2	47.6	47.6	47.6	44.5	42.3
0.6	0.8*	2.6	2.3*	2.6	3.0^	3.0^	3.0^	2.5	2.2*	34.7	52.0	49.3	52.0	54.8	54.8	54.8	51.3	48.6
0.8	0.2*	3.2^	2.3*	3.2^	3.4^	3.4^	3.4^	2.9^	2.4	31.7	66.6	60.9	66.6	67.9	67.9	67.9	64.8	61.5

Table 1.2. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k=1, n_k~constant, p_C=0.2, δ_0=-0.05

Correlation Structure	***Type I error rate (%)***									***Power (%)***								
r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	2.4	2.3*	2.3*	2.3*	2.5	2.5	2.5	2.3*	2.2*	24.0	23.9	23.8	23.8	24.4	24.4	24.4	23.8	22.9
0.2	1.9*	2.5	2.5	2.5	2.6	2.6	2.6	2.5	2.3*	22.3	26.0	25.6	25.6	26.2	26.2	26.2	25.7	24.6
0.4	1.3*	2.7^	2.5	2.5	2.7^	2.7^	2.7^	2.6	2.4	20.4	29.4	28.3	28.3	29.5	29.5	29.5	29.1	27.7
0.6	0.7*	2.7^	2.4	2.5	2.7^	2.7^	2.7^	2.7^	2.4	18.1	34.9	33.5	34.2	35.2	35.2	35.2	34.9	33.3
0.8	0.2*	2.7^	2.5	2.7^	3.1^	3.1^	3.1^	2.8^	2.5	14.6	46.2	44.8	46.2	48.8	48.8	48.8	46.7	44.7

Table 1.3. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k=1, n_k~constant, p_C=0.3, δ_0=-0.05

Correlation Structure	***Type I error rate (%)***									***Power (%)***								
r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	2.5	2.4	2.5	2.4	2.6	2.6	2.6	2.5	2.3*	19.4	18.9	19.0	18.9	19.5	19.5	19.5	19.1	18.2
0.2	1.9*	2.5	2.5	2.5	2.7^	2.7^	2.7^	2.6	2.4	17.9	21.0	21.0	21.0	21.7	21.7	21.7	21.2	20.2
0.4	1.2*	2.5	2.4	2.4	2.6	2.6	2.6	2.5	2.3*	15.8	23.7	23.7	23.7	24.3	24.3	24.3	24.1	22.8
0.6	0.6*	2.5	2.3*	2.3*	2.5	2.5	2.5	2.5	2.3*	13.8	29.5	28.7	28.7	29.5	29.5	29.5	29.5	28.2
0.8	0.2*	2.6	2.4	2.6	2.8^	2.8^	2.8^	2.8^	2.4	9.3	38.5	37.2	38.3	39.6	39.6	39.6	39.6	37.3

Table 1.4. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=1, n_k~constant, p_C=0.1, δ_0=-0.05

Correlation Structure	***Type I error rate (%)***									***Power (%)***								
r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	2.6	2.7^	2.6	2.7^	2.7^	2.7^	2.7^	2.6	2.4	74.1	74.7	74.0	74.7	75.1	75.1	75.1	74.0	73.0
0.2	2.1*	2.7^	2.6	2.7^	2.8^	2.8^	2.8^	2.6	2.5	74.9	78.1	77.3	78.1	78.3	78.3	78.3	77.4	76.6
0.4	1.5*	2.7^	2.6	2.7^	2.7^	2.7^	2.7^	2.6	2.4	76.5	83.2	82.1	83.2	83.2	83.2	83.2	82.5	81.8
0.6	0.8*	2.7^	2.4	2.7^	2.7^	2.7^	2.7^	2.6	2.4	78.6	89.0	87.8	89.0	89.1	89.1	89.1	88.4	87.7
0.8	0.3*	2.9^	2.4	2.9^	3.1^	3.1^	3.1^	2.8^	2.5	82.4	96.1	95.4	96.1	96.4	96.4	96.4	96.0	95.5

Table 1.5. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=1, n_k~constant, p_C=0.2, δ_0=-0.05

Correlation Structure	***Type I error rate (%)***									***Power (%)***								
r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	2.5	2.4	2.4	2.4	2.5	2.5	2.5	2.4	2.4	50.7	50.5	50.5	50.5	51.0	51.0	51.0	50.5	49.9
0.2	2.0*	2.6	2.5	2.6	2.6	2.6	2.6	2.6	2.5	50.7	55.1	54.8	55.1	55.4	55.4	55.4	55.1	54.3
0.4	1.2*	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.4	50.5	60.9	60.5	60.7	61.2	61.2	61.2	60.8	59.9
0.6	0.7*	2.5	2.4	2.5	2.6	2.6	2.6	2.6	2.3*	50.6	69.9	69.4	69.9	70.3	70.3	70.3	70.0	68.9
0.8	0.1*	2.6	2.4	2.6	2.6	2.6	2.6	2.6	2.4	50.8	84.2	83.1	84.2	84.2	84.2	84.2	84.2	83.1

Table 1.6. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=1, n_k~constant, p_C=0.3, δ_0=-0.05

Correlation Structure	Type I error rate (%)									Power (%)								
r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	2.5	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.3*	40.8	40.4	40.7	40.3	40.7	40.7	40.7	40.7	39.9
0.2	1.8*	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.3*	39.9	44.6	44.7	44.4	44.8	44.8	44.8	44.8	44.0
0.4	1.2*	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.4	39.9	51.4	51.4	51.4	51.8	51.8	51.8	51.8	50.9
0.6	0.6*	2.6	2.5	2.6	2.6	2.6	2.6	2.6	2.5	38.4	59.7	59.3	59.5	59.8	59.8	59.8	59.8	58.7
0.8	0.1*	2.6	2.5	2.6	2.7^	2.7^	2.7^	2.7^	2.5	36.3	75.6	74.9	75.6	75.9	75.9	75.9	75.8	74.8

Table 1.7. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k=1, n_k~constant, p_C=0.1, δ_0=-0.05

Correlation Structure	***Type I error rate (%)***									***Power (%)***								
r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	2.5	2.6	2.5	2.6	2.6	2.6	2.6	2.5	2.4	96.0	96.1	95.9	96.1	96.1	96.1	96.1	95.9	95.8
0.2	1.9*	2.5	2.4	2.5	2.5	2.5	2.5	2.4	2.3*	96.6	97.3	97.2	97.3	97.3	97.3	97.3	97.2	97.0
0.4	1.5*	2.6	2.5	2.6	2.7^	2.7^	2.7^	2.6	2.5	97.3	98.5	98.4	98.5	98.5	98.5	98.5	98.4	98.3
0.6	0.8*	2.6	2.4	2.6	2.6	2.6	2.6	2.5	2.4	98.2	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4
0.8	0.2*	2.7^	2.4	2.7^	2.7^	2.7^	2.7^	2.7^	2.4	99.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.8. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k=1, n_k~constant, p_C=0.2, δ_0=-0.05

Correlation Structure	***Type I error rate (%)***									***Power (%)***								
r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	2.5	2.5	2.5	2.5	2.6	2.6	2.6	2.5	2.5	79.8	79.9	79.8	79.8	80.0	80.0	80.0	79.8	79.4
0.2	1.9*	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.4	81.2	84.1	84.1	84.1	84.3	84.3	84.3	84.1	83.7
0.4	1.3*	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.3*	83.0	88.8	88.7	88.8	88.9	88.9	88.9	88.8	88.5
0.6	0.6*	2.4	2.3*	2.4	2.4	2.4	2.4	2.4	2.3*	85.3	94.0	93.8	94.0	94.0	94.0	94.0	94.0	93.7
0.8	0.2*	2.6	2.5	2.6	2.7^	2.7^	2.7^	2.7^	2.5	89.4	98.7	98.5	98.6	98.7	98.7	98.7	98.7	98.5

Table 1.9. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k=1, n_k~constant, p_C=0.3, δ_0=-0.05

Correlation Structure	***Type I error rate (%)***									***Power (%)***								
r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	2.5	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	68.7	68.6	68.7	68.6	68.8	68.8	68.8	68.7	68.3
0.2	1.8*	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	69.7	73.8	73.9	73.8	73.9	73.9	73.9	73.9	73.5
0.4	1.2*	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.4	71.1	80.3	80.2	80.2	80.4	80.4	80.4	80.4	79.9
0.6	0.6*	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	73.3	87.8	87.7	87.8	87.9	87.9	87.9	87.9	87.5
0.8	0.1*	2.6	2.5	2.6	2.6	2.6	2.6	2.6	2.5	76.9	96.1	95.9	96.1	96.2	96.2	96.2	96.2	95.9

Table 1.10. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k=1, n_k~constant, p_C=0.05, δ_0=-0.02

Correlation Structure	***Type I error rate (%)***									***Power (%)***								
r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	2.4	2.5	2.4	2.5	2.6	2.6	2.6	2.4	2.3*	52.6	53.7	52.3	53.5	53.9	53.9	53.9	52.3	51.5
0.2	2.1*	2.5	2.4	2.5	2.6	2.6	2.6	2.4	2.3*	52.6	56.0	54.5	55.6	56.2	56.2	56.2	54.6	53.7
0.4	1.7*	2.7^	2.4	2.6	2.7^	2.7^	2.7^	2.5	2.3*	53.3	60.4	59.0	60.1	60.8	60.8	60.8	59.2	58.1
0.6	1.0*	2.5	2.3*	2.5	2.6	2.6	2.6	2.4	2.2*	53.5	67.2	65.6	67.1	67.5	67.5	67.5	65.9	64.8
0.8	0.4*	2.8^	2.4	2.8^	2.8^	2.8^	2.8^	2.6	2.4	54.0	80.4	78.1	80.4	80.4	80.4	80.4	79.3	78.0

Table 1.11. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=1, n_k~constant, p_C=0.05, δ_0=-0.02

Correlation Structure	***Type I error rate (%)***									***Power (%)***								
r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	2.3*	2.6	2.3*	2.6	2.6	2.6	2.6	2.3*	2.1*	29.2	30.7	29.3	30.7	30.8	30.8	30.8	29.3	27.9
0.2	2.2*	2.7^	2.5	2.7^	2.7^	2.7^	2.7^	2.5	2.3*	29.1	32.8	31.2	32.8	32.8	32.8	32.8	31.2	29.8
0.4	1.6*	2.8^	2.5	2.8^	2.8^	2.8^	2.8^	2.5	2.3*	28.1	35.6	33.6	35.6	35.6	35.6	35.6	33.7	32.0
0.6	1.1*	2.9^	2.5	2.9^	2.9^	2.9^	2.9^	2.6	2.3*	26.2	40.4	37.5	40.4	40.8	40.8	40.8	38.2	36.1
0.8	0.4*	2.9^	2.4	2.9^	3.3^	3.3^	3.3^	2.7^	2.3*	23.4	51.3	47.6	51.3	53.6	53.6	53.6	50.2	47.4

Table 1.12. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=1, n_k~constant, p_C=0.2, δ_0=-0.02

Correlation Structure	*Type I error rate (%)*									*Power (%)*								
r_{CE}	*FM*	*MOM*	*ICC*	*O*	*GEE Ind*	*GEE Exc*	*GEE Un*	*AFM*	*AFM CC*	*FM*	*MOM*	*ICC*	*O*	*GEE Ind*	*GEE Exc*	*GEE Un*	*AFM*	*AFM CC*
0	2.4	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.3*	12.1	12.1	12.1	12.1	12.3	12.3	12.3	12.1	11.6
0.2	1.9*	2.5	2.5	2.5	2.6	2.6	2.6	2.5	2.3*	11.0	13.0	13.0	13.0	13.3	13.3	13.3	13.1	12.6
0.4	1.3*	2.5	2.5	2.5	2.6	2.6	2.6	2.5	2.4	9.1	14.2	14.2	14.2	14.5	14.5	14.5	14.2	13.8
0.6	0.6*	2.4	2.4	2.4	2.5	2.5	2.5	2.4	2.3*	6.9	16.5	16.3	16.5	16.8	16.8	16.8	16.5	15.9
0.8	0.2*	2.5	2.4	2.5	2.6	2.6	2.6	2.5	2.3*	4.0	22.3	21.9	22.3	22.7	22.7	22.7	22.4	21.2

Table 1.13. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=1, n_k~constant, p_C=0.2, δ_0=-0.1

Correlation Structure	***Type I error rate (%)***									***Power (%)***								
r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.3*	97.7	97.7	97.7	97.7	97.8	97.8	97.8	97.7	97.6
0.2	1.9*	2.5	2.5	2.5	2.6	2.6	2.6	2.5	2.4	98.3	98.6	98.6	98.6	98.6	98.6	98.6	98.6	98.5
0.4	1.2*	2.4	2.3*	2.4	2.5	2.5	2.5	2.4	2.3*	98.8	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.3
0.6	0.7*	2.5	2.3*	2.5	2.5	2.5	2.5	2.5	2.4	99.4	99.9	99.8	99.9	99.9	99.9	99.9	99.9	99.8
0.8	0.2*	2.8^	2.4	2.7^	2.9^	2.9^	2.9^	2.9^	2.6	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.14. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=1, n_k~constant, p_C=0.3, δ_0=-0.1

Correlation Structure	Type I error rate (%)									Power (%)								
r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	2.4	2.4	2.5	2.4	2.5	2.5	2.5	2.5	2.4	93.5	93.3	93.4	93.3	93.4	93.4	93.4	93.4	93.2
0.2	1.8*	2.5	2.5	2.5	2.5	2.5	2.5	2.6	2.4	94.4	95.6	95.6	95.5	95.6	95.6	95.6	95.7	95.4
0.4	1.1*	2.3*	2.4	2.3*	2.4	2.4	2.4	2.4	2.3*	95.9	97.8	97.8	97.8	97.9	97.9	97.9	97.9	97.7
0.6	0.5*	2.5	2.4	2.5	2.5	2.5	2.5	2.6	2.4	97.3	99.3	99.2	99.3	99.3	99.3	99.3	99.3	99.2
0.8	0.1*	2.6	2.3*	2.6	2.7^	2.7^	2.7^	2.7^	2.5	98.9	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.15. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=1, n_k~constant, p_C=0.5, δ_0=-0.1

Correlation Structure	Type I error rate (%)									Power (%)								
r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	2.5	2.4	2.4	2.4	2.4	2.4	2.4	2.5	2.4	89.3	88.5	88.8	88.5	88.7	88.7	88.7	88.8	88.5
0.2	1.9*	2.4	2.5	2.4	2.5	2.5	2.5	2.6	2.4	90.7	92.2	92.3	92.2	92.3	92.3	92.3	92.4	92.2
0.4	1.2*	2.5	2.5	2.4	2.5	2.5	2.5	2.6	2.4	92.4	95.7	95.7	95.6	95.7	95.7	95.7	95.8	95.6
0.6	0.6*	2.5	2.4	2.4	2.6	2.6	2.6	2.6	2.4	94.8	98.6	98.5	98.5	98.6	98.6	98.6	98.7	98.5
0.8	0.1*	2.5	2.3*	2.5	2.5	2.5	2.5	2.6	2.4	97.2	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8

Table 1.16. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k=2, n_k~constant, p_C=0.1, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.3*	3.2^	3.1^	3.1^	3.3^	3.3^	3.4^	2.5	2.3*	42.7	48.6	47.7	48.4	49.5	49.5	49.5	44.0	42.8
0.2	0	2.3*	3.0^	2.9^	3.0^	3.1^	3.1^	3.1^	2.4	2.2*	42.9	47.8	46.8	47.6	48.6	48.6	48.7	43.1	41.9
0.4	0	2.5	2.9^	2.8^	2.9^	3.1^	3.1^	3.1^	2.4	2.2*	43.2	46.9	45.7	46.7	47.7	47.7	47.7	42.2	40.9
0.6	0	2.7^	2.9^	2.7^	2.8^	3.0^	3.0^	3.0^	2.3*	2.1*	43.1	45.4	44.0	45.1	46.1	46.1	46.2	40.6	39.1
0.8	0	3.0^	2.8^	2.7^	2.8^	3.0^	3.0^	3.0^	2.3*	2.1*	43.0	43.0	41.7	42.8	43.8	43.8	43.9	38.4	36.9
0	0.2	1.7*	3.1^	2.9^	3.1^	3.2^	3.2^	3.4^	2.3*	2.2*	42.3	53.0	51.6	52.8	53.8	53.8	54.3	47.8	46.4
0.2	0.2	1.9*	3.2^	3.1^	3.2^	3.4^	3.4^	3.5^	2.6	2.3*	42.6	52.0	50.6	51.7	52.8	52.8	53.2	46.8	45.4
0.4	0.2	1.9*	2.9^	2.8^	2.9^	3.1^	3.1^	3.1^	2.3*	2.1*	42.5	50.5	49.0	50.3	51.3	51.3	51.7	45.3	43.8
0.6	0.2	2.1*	2.9^	2.7^	2.8^	3.0^	3.0^	3.1^	2.3*	2.1*	42.7	48.9	47.3	48.7	49.7	49.7	50.1	43.9	42.4
0.8	0.2	2.4	2.9^	2.7^	2.8^	3.0^	3.0^	3.0^	2.3*	2.1*	42.4	46.2	44.6	46.0	47.0	47.0	47.3	41.3	39.6
0.4	0.4	1.4*	3.2^	2.9^	3.2^	3.4^	3.4^	3.6^	2.4	2.2*	41.7	56.2	54.4	56.1	57.0	57.0	58.0	51.0	49.1
0.6	0.4	1.5*	3.1^	2.8^	3.1^	3.2^	3.2^	3.5^	2.4	2.2*	42.0	54.7	52.6	54.5	55.5	55.5	56.2	49.2	47.3
0	0.5	0.9*	3.4^	3.1^	3.4^	3.6^	3.6^	3.9^	2.5	2.3*	41.3	64.6	62.6	64.5	65.4	65.4	66.7	59.0	57.1
0.8	0.5	1.3*	3.1^	2.7^	3.0^	3.2^	3.2^	3.5^	2.3*	2.0*	42.2	55.9	53.5	55.7	56.7	56.7	57.8	50.5	48.4
0.6	0.6	0.8*	3.2^	2.8^	3.2^	3.4^	3.4^	3.8^	2.4	2.1*	41.3	64.2	61.4	64.1	64.9	64.9	66.4	58.5	56.1
0.8	0.8	0.2*	3.6^	2.7^	3.5^	3.6^	3.6^	4.5^	2.6	2.1*	39.7	77.6	73.5	77.3	77.8	77.8	80.0	72.1	69.1

Table 1.17. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k=2, n_k~constant, p_C=0.2, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.8^	2.8^	2.7^	2.9^	2.9^	2.9^	2.5	2.3*	28.6	31.0	31.0	30.7	31.5	31.5	31.7	29.1	28.3
0.2	0	2.6	2.7^	2.7^	2.7^	2.9^	2.9^	2.9^	2.5	2.3*	29.2	30.5	30.5	30.2	31.2	31.2	31.2	28.7	27.9
0.4	0	2.9^	2.7^	2.7^	2.7^	2.9^	2.9^	2.9^	2.5	2.3*	28.9	28.8	28.8	28.6	29.4	29.4	29.4	27.1	26.2
0.6	0	3.0^	2.7^	2.7^	2.7^	2.8^	2.8^	2.8^	2.4	2.3*	29.4	27.8	27.8	27.5	28.4	28.4	28.4	26.0	25.1
0.8	0	3.3^	2.6	2.6	2.5	2.7^	2.7^	2.7^	2.3*	2.1*	29.9	26.5	26.5	26.3	27.1	27.1	27.2	24.9	24.0
0	0.2	1.7*	2.9^	2.9^	2.9^	3.0^	3.0^	3.1^	2.6	2.4	27.1	35.0	35.0	34.7	35.6	35.6	36.1	32.9	31.9
0.2	0.2	1.7*	2.7^	2.7^	2.7^	2.8^	2.8^	2.9^	2.4	2.3*	27.2	33.7	33.6	33.4	34.3	34.3	34.7	31.6	30.6
0.4	0.2	2.0*	2.7^	2.7^	2.7^	2.8^	2.8^	2.9^	2.4	2.3*	27.7	32.5	32.4	32.2	33.2	33.2	33.6	30.4	29.3
0.6	0.2	2.2*	2.7^	2.7^	2.7^	2.8^	2.8^	2.9^	2.4	2.3*	27.9	30.7	30.6	30.5	31.4	31.4	31.8	29.0	27.9
0.8	0.2	2.5	2.6	2.6	2.6	2.7^	2.7^	2.8^	2.4	2.2*	28.6	29.2	29.0	28.9	29.8	29.8	30.1	27.5	26.4
0.4	0.4	1.2*	2.8^	2.8^	2.8^	2.9^	2.9^	3.1^	2.5	2.3*	25.7	37.7	37.4	37.4	38.4	38.4	39.1	35.5	34.0
0.6	0.4	1.4*	2.7^	2.6	2.6	2.8^	2.8^	2.9^	2.4	2.2*	26.3	35.8	35.5	35.6	36.5	36.5	37.2	33.7	32.3
0	0.5	0.6*	3.0^	2.9^	2.9^	3.1^	3.1^	3.4^	2.6	2.4	23.9	45.6	45.4	45.4	46.4	46.4	47.7	43.3	41.7
0.8	0.5	1.2*	2.6	2.6	2.6	2.8^	2.8^	3.0^	2.3*	2.1*	26.1	37.1	36.5	36.8	37.9	37.9	38.7	35.0	33.4
0.6	0.6	0.7*	2.8^	2.7^	2.8^	3.0^	3.0^	3.2^	2.5	2.2*	23.1	44.2	43.5	44.0	45.0	45.0	46.4	41.8	39.7
0.8	0.8	0.1*	3.0^	2.7^	3.0^	3.2^	3.2^	3.6^	2.6	2.1*	19.8	58.4	56.4	58.3	59.2	59.2	61.0	55.7	52.6

Table 1.18. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k=2, n_k~constant, p_C=0.3, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.5	2.7^	2.7^	2.7^	2.8^	2.8^	2.8^	2.6	2.4	23.6	24.6	24.9	24.4	25.2	25.2	25.2	23.9	23.2
0.2	0	2.6	2.6	2.7^	2.6	2.7^	2.7^	2.7^	2.5	2.4	24.2	24.0	24.3	23.8	24.6	24.6	24.6	23.3	22.7
0.4	0	3.0^	2.7^	2.8^	2.7^	2.8^	2.8^	2.8^	2.6	2.4	24.2	22.7	23.0	22.5	23.2	23.2	23.4	22.1	21.3
0.6	0	3.1^	2.5	2.6	2.5	2.6	2.6	2.7^	2.4	2.3*	24.7	21.8	22.2	21.7	22.4	22.4	22.5	21.2	20.5
0.8	0	3.4^	2.5	2.5	2.5	2.6	2.6	2.6	2.4	2.3*	25.2	20.9	21.1	20.8	21.5	21.5	21.6	20.3	19.6
0	0.2	1.5*	2.5	2.6	2.5	2.6	2.6	2.7^	2.4	2.3*	21.5	28.2	28.3	27.9	28.8	28.8	29.1	27.3	26.5
0.2	0.2	1.8*	2.6	2.7^	2.6	2.8^	2.8^	2.8^	2.5	2.4	22.2	27.0	27.1	26.7	27.6	27.6	27.9	26.2	25.4
0.4	0.2	2.1*	2.6	2.7^	2.6	2.7^	2.7^	2.8^	2.5	2.4	22.7	26.1	26.2	25.8	26.6	26.6	26.9	25.3	24.4
0.6	0.2	2.2*	2.6	2.6	2.5	2.6	2.6	2.7^	2.5	2.3*	22.7	24.5	24.6	24.3	25.0	25.0	25.3	23.8	22.9
0.8	0.2	2.7^	2.7^	2.7^	2.6	2.8^	2.8^	2.9^	2.6	2.4	23.8	23.8	23.9	23.6	24.3	24.3	24.6	23.1	22.2
0.4	0.4	1.2*	2.7^	2.7^	2.6	2.8^	2.8^	2.9^	2.6	2.4	19.9	30.5	30.6	30.2	31.2	31.2	31.7	29.6	28.4
0.6	0.4	1.4*	2.6	2.6	2.6	2.7^	2.7^	2.8^	2.5	2.3*	20.9	29.1	29.1	28.9	29.8	29.8	30.1	28.3	27.0
0	0.5	0.5*	2.6	2.6	2.5	2.7^	2.7^	2.9^	2.5	2.3*	17.2	38.3	38.3	38.0	39.0	39.0	39.8	37.3	35.8
0.8	0.5	1.2*	2.7^	2.7^	2.7^	2.8^	2.8^	2.9^	2.5	2.3*	20.5	30.0	29.9	29.8	30.7	30.7	31.2	29.3	27.9
0.6	0.6	0.5*	2.7^	2.6	2.6	2.8^	2.8^	3.0^	2.5	2.2*	17.4	36.2	36.0	35.9	36.9	36.9	37.7	35.2	33.5
0.8	0.8	0.1*	2.6	2.4	2.6	2.8^	2.8^	3.1^	2.5	2.1*	13.2	49.2	48.0	49.0	50.0	50.0	51.3	48.2	45.5

Table 1.19. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=2, n_k~constant, p_C=0.1, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.3*	2.9^	2.8^	2.9^	3.0^	3.0^	3.0^	2.5	2.4	82.1	84.4	84.0	84.4	84.6	84.6	84.7	82.7	82.2
0.2	0	2.4	2.8^	2.8^	2.8^	2.9^	2.9^	2.9^	2.4	2.3*	82.1	83.9	83.5	83.9	84.1	84.1	84.1	82.1	81.6
0.4	0	2.6	2.8^	2.7^	2.8^	2.9^	2.9^	2.9^	2.5	2.4	81.8	82.9	82.4	82.9	83.1	83.1	83.1	81.0	80.4
0.6	0	2.8^	2.7^	2.6	2.7^	2.7^	2.7^	2.8^	2.4	2.3*	81.4	81.5	81.0	81.5	81.8	81.8	81.9	79.5	78.9
0.8	0	3.2^	2.8^	2.7^	2.8^	2.8^	2.8^	2.9^	2.4	2.3*	80.6	79.5	78.9	79.5	79.7	79.7	79.7	77.3	76.5
0	0.2	1.7*	2.8^	2.7^	2.8^	2.9^	2.9^	2.9^	2.4	2.3*	83.7	88.2	87.8	88.2	88.4	88.4	88.5	86.6	86.2
0.2	0.2	1.9*	2.9^	2.8^	2.9^	2.9^	2.9^	3.0^	2.5	2.4	83.4	87.5	87.1	87.5	87.7	87.7	87.8	85.8	85.3
0.4	0.2	2.0*	2.8^	2.7^	2.8^	2.9^	2.9^	3.0^	2.4	2.3*	82.7	86.2	85.7	86.2	86.4	86.4	86.5	84.5	83.9
0.6	0.2	2.2*	2.7^	2.6	2.7^	2.8^	2.8^	2.8^	2.3*	2.2*	82.4	85.1	84.5	85.1	85.3	85.3	85.5	83.2	82.6
0.8	0.2	2.5	2.7^	2.6	2.7^	2.7^	2.7^	2.8^	2.3*	2.2*	81.9	83.3	82.6	83.3	83.5	83.5	83.6	81.3	80.5
0.4	0.4	1.4*	2.9^	2.7^	2.9^	2.9^	2.9^	3.0^	2.4	2.3*	84.9	91.0	90.5	90.9	91.1	91.1	91.4	89.6	89.1
0.6	0.4	1.5*	2.8^	2.7^	2.8^	2.9^	2.9^	3.0^	2.4	2.3*	84.2	89.7	89.2	89.7	89.8	89.8	90.1	88.2	87.6
0	0.5	0.9*	3.0^	2.9^	3.0^	3.1^	3.1^	3.3^	2.6	2.4	86.8	95.1	94.7	95.1	95.2	95.2	95.4	94.1	93.7
0.8	0.5	1.4*	2.8^	2.6	2.8^	2.9^	2.9^	3.0^	2.4	2.3*	84.7	90.6	90.0	90.6	90.8	90.8	91.0	89.2	88.6
0.6	0.6	0.8*	2.9^	2.7^	2.9^	3.0^	3.0^	3.3^	2.5	2.3*	87.1	95.2	94.7	95.1	95.2	95.2	95.5	94.1	93.7
0.8	0.8	0.3*	3.2^	2.6	3.1^	3.2^	3.2^	3.6^	2.5	2.2*	90.5	98.8	98.6	98.8	98.9	98.9	99.0	98.5	98.2

Table 1.20. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=2, n_k~constant, p_C=0.2, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.8^	2.8^	2.7^	2.8^	2.8^	2.8^	2.6	2.5	60.8	62.6	62.6	62.5	62.9	62.9	63.0	61.3	60.8
0.2	0	2.6	2.7^	2.7^	2.6	2.7^	2.7^	2.7^	2.5	2.4	60.7	61.1	61.1	61.0	61.4	61.4	61.3	59.8	59.2
0.4	0	2.8^	2.7^	2.7^	2.7^	2.7^	2.7^	2.7^	2.5	2.4	60.3	59.6	59.6	59.5	59.9	59.9	59.9	58.3	57.6
0.6	0	3.0^	2.6	2.6	2.6	2.6	2.6	2.6	2.4	2.3*	59.9	57.6	57.6	57.5	57.9	57.9	57.9	56.4	55.6
0.8	0	3.3^	2.5	2.5	2.5	2.6	2.6	2.6	2.4	2.3*	59.8	55.3	55.3	55.2	55.6	55.6	55.6	54.1	53.3
0	0.2	1.7*	2.8^	2.8^	2.8^	2.8^	2.8^	2.8^	2.6	2.4	60.8	68.0	68.0	68.0	68.3	68.3	68.6	66.9	66.3
0.2	0.2	1.7*	2.5	2.5	2.5	2.6	2.6	2.6	2.3*	2.2*	61.1	67.0	67.0	67.0	67.3	67.3	67.4	65.8	65.0
0.4	0.2	2.1*	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.5	2.4	61.1	65.4	65.3	65.3	65.7	65.7	65.8	64.2	63.4
0.6	0.2	2.2*	2.6	2.6	2.5	2.6	2.6	2.7^	2.4	2.3*	60.5	62.9	62.8	62.8	63.2	63.2	63.4	61.7	61.0
0.8	0.2	2.6	2.6	2.6	2.6	2.6	2.6	2.7^	2.4	2.3*	60.0	60.2	60.1	60.1	60.5	60.5	60.8	58.9	58.1
0.4	0.4	1.1*	2.6	2.6	2.6	2.7^	2.7^	2.8^	2.4	2.3*	61.6	72.9	72.7	72.9	73.1	73.1	73.5	71.6	70.8
0.6	0.4	1.4*	2.6	2.5	2.5	2.6	2.6	2.7^	2.4	2.3*	61.3	70.6	70.3	70.5	70.8	70.8	71.2	69.4	68.5
0	0.5	0.6*	2.6	2.6	2.6	2.7^	2.7^	2.8^	2.4	2.3*	63.0	82.5	82.3	82.5	82.7	82.7	83.2	81.5	80.8
0.8	0.5	1.2*	2.7^	2.7^	2.7^	2.8^	2.8^	2.8^	2.5	2.3*	61.9	72.6	72.3	72.5	72.8	72.8	73.3	71.4	70.4
0.6	0.6	0.6*	2.7^	2.6	2.7^	2.7^	2.7^	2.9^	2.5	2.3*	63.3	81.4	81.0	81.4	81.6	81.6	82.1	80.3	79.6
0.8	0.8	0.1*	2.7^	2.5	2.7^	2.8^	2.8^	3.0^	2.5	2.2*	65.5	92.5	92.0	92.5	92.7	92.7	93.0	91.9	91.2

Table 1.21. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=2, n_k~constant, p_C=0.3, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.5	2.6	2.7^	2.6	2.7^	2.7^	2.7^	2.6	2.5	50.4	51.2	51.5	51.1	51.6	51.6	51.6	50.6	50.1
0.2	0	2.7^	2.5	2.6	2.5	2.6	2.6	2.6	2.5	2.4	50.5	49.8	50.0	49.7	50.1	50.1	50.1	49.2	48.7
0.4	0	2.8^	2.4	2.4	2.4	2.4	2.4	2.4	2.3*	2.2*	50.0	47.7	48.0	47.7	48.1	48.1	48.1	47.2	46.6
0.6	0	3.1^	2.5	2.5	2.5	2.6	2.6	2.6	2.4	2.4	50.1	46.2	46.4	46.1	46.5	46.5	46.6	45.6	45.0
0.8	0	3.4^	2.4	2.4	2.4	2.5	2.5	2.5	2.3*	2.3*	49.9	44.1	44.2	44.0	44.4	44.4	44.4	43.6	42.9
0	0.2	1.5*	2.5	2.5	2.5	2.5	2.5	2.6	2.4	2.3*	50.6	58.5	58.6	58.3	58.7	58.7	58.8	57.9	57.2
0.2	0.2	2.0*	2.7^	2.7^	2.7^	2.7^	2.7^	2.8^	2.6	2.6	50.9	56.7	56.9	56.6	57.0	57.0	57.2	56.2	55.5
0.4	0.2	2.0*	2.5	2.6	2.5	2.6	2.6	2.6	2.5	2.4	50.3	54.2	54.4	54.2	54.6	54.6	54.8	53.7	53.0
0.6	0.2	2.3*	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.5	2.4	50.3	52.2	52.4	52.2	52.5	52.5	52.7	51.7	51.0
0.8	0.2	2.5	2.5	2.5	2.5	2.5	2.5	2.6	2.4	2.3*	50.2	49.7	49.9	49.7	50.1	50.1	50.2	49.2	48.5
0.4	0.4	1.2*	2.5	2.5	2.5	2.6	2.6	2.7^	2.5	2.3*	50.8	63.4	63.4	63.3	63.7	63.7	63.9	62.9	62.0
0.6	0.4	1.4*	2.6	2.6	2.6	2.6	2.6	2.7^	2.5	2.3*	50.1	60.1	60.1	60.0	60.4	60.4	60.7	59.5	58.7
0	0.5	0.5*	2.6	2.6	2.6	2.6	2.6	2.7^	2.5	2.4	50.0	74.2	74.2	74.2	74.4	74.4	74.8	73.6	72.8
0.8	0.5	1.3*	2.7^	2.6	2.6	2.7^	2.7^	2.8^	2.6	2.4	50.7	62.3	62.2	62.2	62.6	62.6	62.9	61.7	60.8
0.6	0.6	0.6*	2.5	2.5	2.5	2.6	2.6	2.7^	2.5	2.3*	50.4	72.0	71.9	72.0	72.3	72.3	72.7	71.5	70.4
0.8	0.8	0.1*	2.5	2.4	2.5	2.6	2.6	2.8^	2.5	2.2*	50.3	86.0	85.7	86.0	86.2	86.2	86.7	85.7	84.9

Table 1.22. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k=2, n_k~constant, p_C=0.1, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.3*	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.4	2.3*	98.5	98.7	98.7	98.7	98.7	98.7	98.7	98.5	98.5
0.2	0	2.6	2.8^	2.7^	2.8^	2.8^	2.8^	2.8^	2.5	2.5	98.4	98.5	98.5	98.5	98.6	98.6	98.6	98.3	98.3
0.4	0	2.8^	2.8^	2.7^	2.8^	2.8^	2.8^	2.8^	2.5	2.4	98.2	98.3	98.2	98.3	98.3	98.3	98.3	98.1	98.0
0.6	0	3.0^	2.7^	2.7^	2.7^	2.8^	2.8^	2.8^	2.5	2.4	98.1	98.1	98.0	98.1	98.1	98.1	98.1	97.8	97.7
0.8	0	3.2^	2.6	2.6	2.6	2.6	2.6	2.6	2.4	2.3*	97.9	97.6	97.5	97.6	97.7	97.7	97.6	97.3	97.2
0	0.2	1.9*	2.9^	2.8^	2.9^	2.9^	2.9^	2.9^	2.6	2.5	98.9	99.3	99.3	99.3	99.3	99.3	99.3	99.2	99.2
0.2	0.2	1.9*	2.8^	2.7^	2.8^	2.8^	2.8^	2.8^	2.5	2.4	98.8	99.2	99.2	99.2	99.2	99.2	99.2	99.1	99.0
0.4	0.2	2.0*	2.7^	2.6	2.7^	2.7^	2.7^	2.7^	2.4	2.3*	98.7	99.1	99.0	99.1	99.1	99.1	99.1	98.9	98.9
0.6	0.2	2.3*	2.7^	2.6	2.7^	2.7^	2.7^	2.8^	2.4	2.3*	98.5	98.8	98.7	98.8	98.8	98.8	98.8	98.6	98.5
0.8	0.2	2.6	2.7^	2.6	2.7^	2.7^	2.7^	2.8^	2.4	2.4	98.3	98.5	98.4	98.5	98.5	98.5	98.5	98.2	98.1
0.4	0.4	1.4*	2.8^	2.7^	2.8^	2.9^	2.9^	3.0^	2.6	2.4	99.2	99.7	99.7	99.7	99.7	99.7	99.7	99.6	99.6
0.6	0.4	1.5*	2.6	2.5	2.6	2.7^	2.7^	2.7^	2.3*	2.2*	99.0	99.5	99.5	99.5	99.5	99.5	99.5	99.4	99.4
0	0.5	0.9*	2.8^	2.7^	2.8^	2.9^	2.9^	3.0^	2.5	2.4	99.6	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
0.8	0.5	1.4*	2.7^	2.6	2.7^	2.8^	2.8^	2.9^	2.4	2.3*	99.1	99.6	99.6	99.6	99.6	99.6	99.6	99.5	99.5
0.6	0.6	0.8*	2.8^	2.6	2.8^	2.9^	2.9^	3.0^	2.5	2.3*	99.6	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
0.8	0.8	0.3*	3.0^	2.6	3.0^	3.0^	3.0^	3.3^	2.6	2.4	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.23. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k=2, n_k~constant, p_C=0.2, δ_0=-0.05

Correlation Structure		**Type I error rate (%)**									**Power (%)**								
r_C	r_{CE}	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**
0	0	2.5	2.7^	2.7^	2.7^	2.7^	2.7^	2.7^	2.5	2.5	88.5	89.2	89.2	89.2	89.2	89.2	89.2	88.7	88.5
0.2	0	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.4	2.3*	88.4	88.3	88.4	88.3	88.4	88.4	88.4	87.9	87.6
0.4	0	2.9^	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.5	2.5	87.7	87.0	87.0	87.0	87.1	87.1	87.1	86.5	86.2
0.6	0	3.1^	2.6	2.6	2.6	2.6	2.6	2.7^	2.5	2.5	87.2	85.7	85.7	85.7	85.8	85.8	85.8	85.3	85.0
0.8	0	3.4^	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.4	86.7	84.0	84.0	84.0	84.1	84.1	84.1	83.5	83.1
0	0.2	1.8*	2.8^	2.7^	2.7^	2.8^	2.8^	2.8^	2.6	2.6	90.2	93.0	93.0	93.0	93.0	93.0	93.1	92.7	92.5
0.2	0.2	1.7*	2.5	2.5	2.5	2.5	2.5	2.6	2.4	2.3*	90.0	92.3	92.3	92.3	92.4	92.4	92.4	92.0	91.8
0.4	0.2	2.1*	2.5	2.5	2.5	2.6	2.6	2.6	2.4	2.3*	89.5	91.3	91.3	91.3	91.4	91.4	91.4	90.9	90.7
0.6	0.2	2.3*	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.4	89.0	90.0	90.0	90.0	90.1	90.1	90.1	89.6	89.3
0.8	0.2	2.5	2.4	2.4	2.4	2.4	2.4	2.5	2.3*	2.2*	88.1	88.1	88.0	88.1	88.1	88.1	88.2	87.6	87.3
0.4	0.4	1.2*	2.6	2.6	2.6	2.6	2.6	2.7^	2.5	2.4	91.5	95.3	95.2	95.3	95.3	95.3	95.4	95.0	94.8
0.6	0.4	1.4*	2.6	2.6	2.6	2.6	2.6	2.7^	2.4	2.3*	90.9	94.3	94.2	94.3	94.3	94.3	94.4	94.0	93.8
0	0.5	0.6*	2.8^	2.7^	2.7^	2.8^	2.8^	2.9^	2.6	2.5	93.7	98.3	98.3	98.3	98.3	98.3	98.4	98.2	98.1
0.8	0.5	1.2*	2.5	2.5	2.5	2.5	2.5	2.6	2.4	2.3*	91.2	94.9	94.8	94.9	95.0	95.0	95.1	94.6	94.4
0.6	0.6	0.6*	2.5	2.5	2.5	2.6	2.6	2.6	2.4	2.2*	93.3	98.0	98.0	98.0	98.1	98.1	98.1	97.9	97.8
0.8	0.8	0.1*	2.7^	2.5	2.7^	2.7^	2.7^	2.9^	2.5	2.3*	96.4	99.7	99.7	99.7	99.8	99.8	99.8	99.7	99.7

Table 1.24. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k=2, n_k~constant, p_C=0.3, δ_0=-0.05

Correlation Structure		**Type I error rate (%)**									**Power (%)**								
r_C	r_{CE}	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**
0	0	2.4	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.4	79.8	80.3	80.4	80.2	80.4	80.4	80.4	79.9	79.7
0.2	0	2.7^	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.5	79.3	78.7	78.8	78.7	78.8	78.8	78.8	78.4	78.1
0.4	0	2.8^	2.4	2.4	2.4	2.4	2.4	2.5	2.4	2.3*	78.7	76.9	77.0	76.9	77.1	77.1	77.1	76.6	76.3
0.6	0	3.1^	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.3*	78.2	75.2	75.3	75.1	75.3	75.3	75.3	74.9	74.5
0.8	0	3.3^	2.4	2.4	2.4	2.4	2.4	2.4	2.3*	2.3*	78.0	73.4	73.5	73.3	73.5	73.5	73.5	73.0	72.7
0	0.2	1.6*	2.5	2.5	2.5	2.5	2.5	2.5	2.4	2.4	81.7	86.3	86.3	86.3	86.4	86.4	86.4	86.0	85.8
0.2	0.2	1.9*	2.6	2.6	2.6	2.6	2.6	2.7^	2.5	2.5	81.3	84.9	85.0	84.9	85.0	85.0	85.1	84.7	84.4
0.4	0.2	2.0*	2.5	2.6	2.5	2.6	2.6	2.6	2.5	2.4	80.4	83.0	83.0	82.9	83.1	83.1	83.1	82.7	82.4
0.6	0.2	2.4	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	80.3	81.5	81.5	81.4	81.6	81.6	81.6	81.2	80.8
0.8	0.2	2.7^	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.5	79.5	79.1	79.1	79.0	79.2	79.2	79.3	78.8	78.4
0.4	0.4	1.1*	2.5	2.5	2.5	2.5	2.5	2.6	2.5	2.4	82.9	89.7	89.7	89.7	89.8	89.8	89.9	89.5	89.2
0.6	0.4	1.4*	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.6	2.5	82.3	88.0	88.0	88.0	88.1	88.1	88.2	87.8	87.5
0	0.5	0.5*	2.6	2.6	2.6	2.6	2.6	2.7^	2.5	2.4	85.9	95.6	95.6	95.6	95.7	95.7	95.7	95.5	95.4
0.8	0.5	1.2*	2.6	2.6	2.6	2.6	2.6	2.7^	2.5	2.4	83.0	89.4	89.4	89.4	89.5	89.5	89.7	89.2	88.9
0.6	0.6	0.6*	2.6	2.6	2.6	2.6	2.6	2.7^	2.6	2.4	85.8	95.0	94.9	94.9	95.0	95.0	95.0	94.8	94.6
0.8	0.8	0.2*	2.6	2.5	2.6	2.6	2.6	2.7^	2.6	2.4	89.9	99.1	99.0	99.1	99.1	99.1	99.1	99.0	98.9

Table 1.25. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k=2, n_k~constant, p_C=0.05, δ_0=-0.02

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.2*	2.9^	2.7^	2.9^	2.9^	2.9^	2.9^	2.3*	2.2*	61.3	65.2	64.4	65.2	65.4	65.4	65.4	62.0	61.3
0.2	0	2.4	2.9^	2.8^	2.9^	3.0^	3.0^	3.0^	2.4	2.3*	60.9	64.4	63.5	64.4	64.6	64.6	64.6	61.1	60.4
0.4	0	2.7^	3.1^	3.0^	3.1^	3.1^	3.1^	3.1^	2.5	2.4	60.9	63.4	62.4	63.4	63.6	63.6	63.6	60.1	59.3
0.6	0	2.7^	2.9^	2.7^	2.9^	2.9^	2.9^	2.9^	2.4	2.3*	60.8	62.0	61.0	61.9	62.1	62.1	62.1	58.6	57.8
0.8	0	3.1^	2.8^	2.7^	2.8^	2.9^	2.9^	2.9^	2.4	2.3*	60.5	59.5	58.6	59.4	59.6	59.6	59.7	56.2	55.2
0	0.2	1.8*	2.8^	2.7^	2.8^	2.9^	2.9^	2.9^	2.3*	2.2*	61.9	68.8	68.0	68.8	69.0	69.0	69.3	65.6	64.8
0.2	0.2	2.1*	3.0^	2.9^	3.0^	3.0^	3.0^	3.0^	2.5	2.4	61.2	67.7	66.8	67.7	67.9	67.9	68.1	64.3	63.5
0.4	0.2	2.1*	3.0^	2.8^	2.9^	3.0^	3.0^	3.0^	2.4	2.3*	61.2	66.7	65.7	66.7	66.9	66.9	67.0	63.3	62.4
0.6	0.2	2.5	3.0^	2.9^	3.0^	3.0^	3.0^	3.1^	2.5	2.4	61.3	65.2	64.1	65.1	65.4	65.4	65.6	61.7	60.9
0.8	0.2	2.6	2.8^	2.7^	2.8^	2.8^	2.8^	2.9^	2.3*	2.2*	61.1	62.8	61.6	62.7	62.9	62.9	63.2	59.4	58.5
0.4	0.4	1.6*	2.9^	2.8^	2.9^	3.0^	3.0^	3.1^	2.4	2.3*	61.8	72.0	71.0	72.0	72.2	72.2	72.6	68.6	67.7
0.6	0.4	1.7*	2.8^	2.7^	2.8^	2.9^	2.9^	2.9^	2.3*	2.2*	61.7	70.0	69.0	70.0	70.2	70.2	70.6	66.6	65.6
0	0.5	1.0*	2.9^	2.7^	2.9^	3.0^	3.0^	3.1^	2.3*	2.2*	62.9	77.8	76.9	77.8	78.0	78.0	78.7	74.6	73.8
0.8	0.5	1.6*	2.9^	2.7^	2.9^	2.9^	2.9^	3.0^	2.4	2.2*	62.0	70.8	69.6	70.8	71.0	71.0	71.5	67.5	66.4
0.6	0.6	1.0*	3.0^	2.8^	3.0^	3.0^	3.0^	3.2^	2.4	2.3*	63.1	78.5	77.3	78.5	78.7	78.7	79.4	75.3	74.3
0.8	0.8	0.3*	3.0^	2.7^	3.0^	3.1^	3.1^	3.5^	2.4	2.2*	64.7	89.0	87.7	89.0	89.1	89.1	89.8	86.5	85.4

Table 1.26. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=2, n_k~constant, p_C=0.05, δ_0=-0.02

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.3*	3.2^	3.1^	3.2^	3.3^	3.3^	3.3^	2.5	2.4	33.6	39.8	38.8	39.8	40.1	40.1	40.2	35.1	34.2
0.2	0	2.4	3.3^	3.1^	3.3^	3.3^	3.3^	3.3^	2.5	2.4	34.0	39.7	38.5	39.7	40.0	40.0	40.0	34.9	33.9
0.4	0	2.5	3.1^	3.0^	3.1^	3.2^	3.2^	3.2^	2.4	2.2*	33.9	38.4	37.1	38.4	38.7	38.7	38.8	33.7	32.7
0.6	0	2.7^	3.2^	3.0^	3.1^	3.2^	3.2^	3.2^	2.4	2.3*	34.5	37.7	36.3	37.7	37.9	37.9	37.9	32.9	31.8
0.8	0	2.8^	2.8^	2.7^	2.8^	2.9^	2.9^	2.9^	2.2*	2.0*	34.6	35.5	34.2	35.5	35.8	35.8	35.9	31.0	29.8
0	0.2	1.8*	3.2^	3.0^	3.2^	3.2^	3.2^	3.3^	2.4	2.3*	33.5	42.8	41.6	42.8	43.2	43.2	43.6	37.9	36.8
0.2	0.2	2.1*	3.4^	3.2^	3.4^	3.5^	3.5^	3.6^	2.6	2.5	33.3	42.0	40.8	42.0	42.3	42.3	42.6	37.0	35.9
0.4	0.2	2.0*	3.0^	2.8^	3.0^	3.1^	3.1^	3.1^	2.3*	2.1*	33.8	41.2	39.8	41.2	41.5	41.5	41.8	36.3	35.1
0.6	0.2	2.1*	2.9^	2.7^	2.9^	3.0^	3.0^	3.1^	2.2*	2.0*	33.8	39.5	38.1	39.5	39.8	39.8	40.2	34.8	33.5
0.8	0.2	2.4	2.9^	2.6	2.9^	2.9^	2.9^	3.0^	2.2*	2.0*	33.7	37.2	35.7	37.2	37.4	37.4	37.7	32.4	31.1
0.4	0.4	1.5*	3.2^	2.9^	3.2^	3.3^	3.3^	3.4^	2.4	2.2*	32.4	45.1	43.3	45.1	45.4	45.4	46.1	39.5	38.2
0.6	0.4	1.8*	3.1^	2.9^	3.1^	3.2^	3.2^	3.4^	2.4	2.2*	33.1	44.0	42.2	44.0	44.3	44.3	45.0	38.5	37.0
0	0.5	1.0*	3.3^	3.0^	3.3^	3.4^	3.4^	3.6^	2.4	2.2*	31.1	50.8	49.0	50.8	51.3	51.3	52.3	44.8	43.2
0.8	0.5	1.6*	3.1^	2.8^	3.1^	3.2^	3.2^	3.4^	2.3*	2.1*	32.5	43.7	41.7	43.7	44.1	44.1	45.0	38.5	36.8
0.6	0.6	0.9*	3.1^	2.8^	3.1^	3.2^	3.2^	3.4^	2.3*	2.1*	31.0	50.9	48.7	50.9	51.4	51.4	52.6	44.9	43.1
0.8	0.8	0.3*	3.3^	2.8^	3.3^	3.4^	3.4^	4.0^	2.3*	1.9*	28.2	63.1	59.5	63.1	63.5	63.5	65.6	56.4	53.5

Table 1.27. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=2, n_k~constant, p_C=0.2, δ_0=-0.02

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.7^	2.7^	2.7^	2.7^	2.7^	2.8^	2.5	2.4	14.1	15.2	15.3	15.2	15.4	15.4	15.4	14.4	14.1
0.2	0	2.7^	2.7^	2.7^	2.7^	2.8^	2.8^	2.8^	2.5	2.5	14.3	14.5	14.5	14.5	14.7	14.7	14.8	13.8	13.4
0.4	0	2.8^	2.7^	2.7^	2.7^	2.7^	2.7^	2.7^	2.5	2.4	15.1	14.4	14.4	14.4	14.5	14.5	14.5	13.7	13.3
0.6	0	3.2^	2.7^	2.7^	2.7^	2.7^	2.7^	2.8^	2.5	2.4	15.8	14.1	14.1	14.1	14.2	14.2	14.3	13.4	13.0
0.8	0	3.4^	2.5	2.5	2.5	2.6	2.6	2.6	2.4	2.3*	16.2	13.4	13.4	13.4	13.6	13.6	13.6	12.8	12.4
0	0.2	1.7*	2.8^	2.8^	2.8^	2.9^	2.9^	2.9^	2.5	2.5	12.5	17.2	17.2	17.2	17.4	17.4	17.6	16.3	15.8
0.2	0.2	1.7*	2.6	2.6	2.6	2.6	2.6	2.7^	2.4	2.3*	12.8	16.4	16.4	16.3	16.5	16.5	16.7	15.4	15.0
0.4	0.2	2.1*	2.7^	2.7^	2.7^	2.7^	2.7^	2.8^	2.5	2.4	13.2	15.6	15.6	15.5	15.7	15.7	15.9	14.8	14.4
0.6	0.2	2.3*	2.7^	2.7^	2.7^	2.7^	2.7^	2.8^	2.5	2.4	13.8	15.1	15.1	15.0	15.2	15.2	15.4	14.3	13.9
0.8	0.2	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.3*	2.2*	14.8	14.7	14.7	14.7	14.9	14.9	15.1	13.9	13.5
0.4	0.4	1.2*	2.5	2.5	2.5	2.6	2.6	2.7^	2.3*	2.2*	11.2	18.1	18.1	18.1	18.3	18.3	18.7	17.1	16.5
0.6	0.4	1.4*	2.6	2.6	2.6	2.7^	2.7^	2.8^	2.4	2.3*	11.8	17.4	17.3	17.3	17.5	17.5	17.9	16.5	15.9
0	0.5	0.6*	2.8^	2.8^	2.8^	2.8^	2.8^	3.0^	2.5	2.3*	8.6	22.0	22.0	22.0	22.3	22.3	22.8	20.7	20.1
0.8	0.5	1.3*	2.6	2.6	2.6	2.6	2.6	2.8^	2.4	2.2*	11.3	17.7	17.6	17.7	17.9	17.9	18.2	16.7	16.1
0.6	0.6	0.7*	2.7^	2.7^	2.7^	2.8^	2.8^	2.9^	2.5	2.3*	8.8	21.4	21.2	21.4	21.6	21.6	22.1	20.2	19.3
0.8	0.8	0.2*	2.7^	2.6	2.7^	2.8^	2.8^	3.0^	2.4	2.2*	5.3	28.3	27.7	28.2	28.6	28.6	29.4	26.6	25.0

Table 1.28. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=2, n_k~constant, p_C=0.2, δ_0=-0.1

Correlation Structure		***Type I error rate (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.7^	2.7^	2.6	2.7^	2.7^	2.7^	2.6	2.5
0.2	0	2.7^	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.5	2.4
0.4	0	2.9^	2.7^	2.7^	2.7^	2.7^	2.7^	2.8^	2.6	2.5
0.6	0	2.9^	2.5	2.5	2.5	2.5	2.5	2.5	2.4	2.3*
0.8	0	3.3^	2.5	2.5	2.5	2.5	2.5	2.5	2.4	2.3*
0	0.2	1.7*	2.6	2.6	2.6	2.6	2.6	2.7^	2.5	2.4
0.2	0.2	1.9*	2.8^	2.8^	2.8^	2.8^	2.8^	2.9^	2.6	2.5
0.4	0.2	1.9*	2.5	2.5	2.5	2.5	2.5	2.6	2.3*	2.2*
0.6	0.2	2.2*	2.5	2.5	2.5	2.6	2.6	2.6	2.4	2.3*
0.8	0.2	2.5	2.6	2.5	2.5	2.6	2.6	2.6	2.4	2.3*
0.4	0.4	1.2*	2.6	2.6	2.6	2.7^	2.7^	2.8^	2.5	2.3*
0.6	0.4	1.4*	2.5	2.5	2.5	2.5	2.5	2.6	2.4	2.2*
0	0.5	0.6*	2.8^	2.7^	2.8^	2.9^	2.9^	3.0^	2.6	2.5
0.8	0.5	1.2*	2.6	2.5	2.6	2.6	2.6	2.7^	2.4	2.3*
0.6	0.6	0.7*	2.8^	2.6	2.7^	2.8^	2.8^	2.9^	2.6	2.5
0.8	0.8	0.2*	2.9^	2.5	2.8^	2.9^	2.9^	3.2^	2.7^	2.4

Power (%)

FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.3	99.3
99.2	99.3	99.3	99.3	99.3	99.3	99.3	99.2	99.2
99.1	99.1	99.1	99.1	99.1	99.1	99.1	99.0	99.0
99.0	98.9	98.9	98.9	98.9	98.9	98.9	98.8	98.7
98.9	98.6	98.6	98.6	98.6	98.6	98.6	98.5	98.4
99.6	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8
99.5	99.7	99.7	99.7	99.7	99.7	99.7	99.7	99.7
99.5	99.6	99.6	99.6	99.6	99.6	99.6	99.6	99.6
99.4	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.4
99.3	99.4	99.3	99.4	99.4	99.4	99.4	99.3	99.2
99.7	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
99.7	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8
99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
99.8	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.29. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=2, n_k~constant, p_C=0.3, δ_0=-0.1

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.4	97.5	97.6	97.6	97.6	97.6	97.6	97.6	97.5	97.5
0.2	0	2.7^	2.5	2.6	2.5	2.6	2.6	2.6	2.5	2.4	97.3	97.2	97.3	97.2	97.3	97.3	97.3	97.2	97.1
0.4	0	2.9^	2.5	2.6	2.5	2.6	2.6	2.6	2.5	2.5	97.2	96.8	96.9	96.8	96.9	96.9	96.9	96.8	96.7
0.6	0	3.1^	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.4	96.8	96.1	96.1	96.0	96.1	96.1	96.1	96.0	95.9
0.8	0	3.6^	2.6	2.7^	2.6	2.7^	2.7^	2.7^	2.6	2.5	96.6	95.4	95.5	95.4	95.5	95.5	95.4	95.3	95.1
0	0.2	1.7*	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.6	2.5	98.4	99.0	99.0	99.0	99.0	99.0	99.0	99.0	98.9
0.2	0.2	1.9*	2.6	2.7^	2.6	2.7^	2.7^	2.7^	2.6	2.5	98.2	98.7	98.7	98.7	98.7	98.7	98.7	98.7	98.6
0.4	0.2	2.0*	2.5	2.6	2.5	2.6	2.6	2.6	2.5	2.4	98.0	98.4	98.4	98.4	98.4	98.4	98.5	98.4	98.3
0.6	0.2	2.1*	2.4	2.4	2.3*	2.4	2.4	2.5	2.4	2.3*	97.7	98.0	98.0	97.9	98.0	98.0	98.0	97.9	97.9
0.8	0.2	2.6	2.5	2.5	2.5	2.5	2.5	2.6	2.5	2.4	97.4	97.4	97.4	97.3	97.4	97.4	97.4	97.3	97.2
0.4	0.4	1.2*	2.6	2.6	2.5	2.6	2.6	2.6	2.5	2.4	98.8	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5
0.6	0.4	1.4*	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.6	2.5	98.7	99.3	99.3	99.3	99.3	99.3	99.3	99.3	99.3
0	0.5	0.5*	2.6	2.6	2.6	2.7^	2.7^	2.8^	2.6	2.5	99.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
0.8	0.5	1.2*	2.5	2.5	2.5	2.6	2.6	2.7^	2.5	2.4	98.7	99.4	99.4	99.4	99.4	99.4	99.5	99.4	99.4
0.6	0.6	0.7*	2.8^	2.7^	2.7^	2.8^	2.8^	2.9^	2.7^	2.6	99.5	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
0.8	0.8	0.2*	2.9^	2.6	2.9^	3.0^	3.0^	3.2^	2.9^	2.6	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.30. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=2, n_k~constant, p_C=0.5, δ_0=-0.1

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.4	2.4	2.4	2.4	2.4	2.4	2.5	2.4	95.6	95.5	95.6	95.5	95.6	95.6	95.6	95.7	95.5
0.2	0	2.8^	2.5	2.5	2.4	2.5	2.5	2.5	2.6	2.5	95.4	94.9	95.0	94.9	95.0	95.0	95.0	95.1	94.9
0.4	0	3.1^	2.4	2.4	2.4	2.4	2.4	2.5	2.5	2.4	94.8	93.7	93.8	93.7	93.8	93.8	93.8	93.9	93.7
0.6	0	3.4^	2.4	2.5	2.4	2.5	2.5	2.5	2.6	2.5	94.4	92.7	92.9	92.7	92.8	92.8	92.8	92.9	92.7
0.8	0	3.9^	2.5	2.6	2.5	2.6	2.6	2.6	2.6	2.5	94.0	91.6	91.8	91.6	91.7	91.7	91.7	91.8	91.6
0	0.2	1.5*	2.5	2.5	2.4	2.5	2.5	2.5	2.6	2.5	96.9	98.0	98.1	98.0	98.1	98.1	98.1	98.1	98.1
0.2	0.2	1.8*	2.5	2.5	2.5	2.5	2.5	2.5	2.6	2.5	96.6	97.4	97.5	97.4	97.5	97.5	97.5	97.5	97.4
0.4	0.2	2.0*	2.4	2.4	2.4	2.4	2.4	2.4	2.5	2.4	96.3	96.8	96.9	96.8	96.9	96.9	96.9	97.0	96.8
0.6	0.2	2.4	2.5	2.5	2.5	2.5	2.5	2.5	2.6	2.5	95.8	96.0	96.1	96.0	96.1	96.1	96.1	96.2	96.0
0.8	0.2	2.8^	2.5	2.5	2.5	2.5	2.5	2.5	2.6	2.5	95.3	94.9	95.1	94.9	95.0	95.0	95.0	95.1	94.9
0.4	0.4	1.1*	2.5	2.5	2.5	2.5	2.5	2.5	2.6	2.5	97.7	98.9	98.9	98.9	99.0	99.0	99.0	99.0	98.9
0.6	0.4	1.4*	2.5	2.5	2.4	2.5	2.5	2.5	2.6	2.5	97.1	98.4	98.4	98.4	98.5	98.5	98.5	98.5	98.4
0	0.5	0.4*	2.4	2.4	2.4	2.5	2.5	2.5	2.6	2.4	98.9	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8
0.8	0.5	1.2*	2.5	2.5	2.4	2.5	2.5	2.5	2.6	2.5	97.5	98.7	98.7	98.7	98.7	98.7	98.8	98.8	98.7
0.6	0.6	0.5*	2.5	2.5	2.5	2.5	2.5	2.6	2.6	2.5	98.6	99.7	99.7	99.7	99.7	99.7	99.7	99.7	99.7
0.8	0.8	0.1*	2.6	2.4	2.6	2.7^	2.7^	2.8^	2.8^	2.6	99.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.31. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k≤2, n_k~uniform, p_C=0.1, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.9^	3.1^	3.2^	3.1^	3.1^	3.2^	2.6	2.4	40.6	43.3	42.2	43.2	45.8	45.9	46.0	42.0	40.6
0.2	0	2.5	2.9^	3.0^	3.1^	3.0^	3.0^	3.1^	2.6	2.4	40.7	43.0	41.5	42.4	45.1	45.1	45.2	41.3	40.0
0.4	0	2.7^	3.0^	3.0^	3.1^	3.1^	3.1^	3.1^	2.6	2.4	41.2	42.5	40.8	41.8	44.5	44.4	44.4	40.7	39.2
0.6	0	2.7^	2.8^	2.7^	2.8^	2.9^	2.9^	[99]	2.4	2.2*	40.9	41.6	39.3	40.4	42.7	42.8	[99]	39.1	37.7
0.8	0	3.1^	2.8^	2.8^	2.9^	2.9^	2.9^	[99]	2.5	2.3*	40.7	40.5	37.5	38.7	40.9	40.9	[99]	37.2	35.5
0	0.2	1.8*	2.8^	2.8^	3.0^	3.0^	3.0^	3.1^	2.5	2.3*	40.4	46.7	45.4	46.7	49.4	49.4	49.8	45.6	44.0
0.2	0.2	2.0*	2.9^	3.0^	3.1^	3.1^	3.1^	3.2^	2.6	2.4	40.4	46.2	44.7	46.0	48.6	48.6	[99]	44.6	43.2
0.4	0.2	2.1*	2.8^	2.9^	3.0^	3.0^	3.0^	3.1^	2.5	2.3*	40.1	45.5	43.2	44.6	47.3	47.2	47.5	43.2	41.6
0.6	0.2	2.1*	2.7^	2.8^	2.9^	2.9^	2.9^	[99]	2.4	2.2*	40.4	44.7	41.9	43.3	46.0	46.1	[99]	42.1	40.4
0.8	0.2	2.5	2.8^	2.8^	3.0^	2.9^	2.9^	[99]	2.5	2.2*	40.7	44.1	40.5	42.1	44.3	44.6	[99]	40.4	38.7
0.4	0.4	1.4*	2.9^	2.8^	3.1^	3.1^	3.1^	3.2^	2.6	2.3*	39.5	50.8	48.4	50.3	52.8	52.9	53.3	48.6	46.7
0.6	0.4	1.5*	2.9^	2.8^	3.0^	3.0^	3.0^	[99]	2.5	2.2*	39.4	49.7	46.7	48.7	51.0	51.3	[99]	46.5	44.6
0	0.5	0.9*	2.8^	2.8^	3.1^	3.1^	3.1^	3.4^	2.6	2.3*	38.8	55.9	54.4	56.6	59.2	59.2	60.4	54.9	52.8
0.8	0.5	1.4*	2.9^	2.7^	3.0^	3.0^	3.0^	[99]	2.5	2.2*	39.9	52.6	47.8	50.5	52.7	53.3	[99]	48.5	46.2
0.6	0.6	0.8*	2.8^	2.6	3.0^	2.9^	2.9^	[99]	2.4	2.1*	38.8	58.1	54.3	57.3	59.5	59.9	[99]	55.0	52.6
0.8	0.8	0.3*	3.4^	2.8^	3.6^	3.4^	3.5^	[99]	2.8^	2.4	36.9	72.2	66.1	70.9	71.9	73.4	[99]	67.5	64.6

Table 1.32. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k≤2, n_k~uniform, p_C=0.2, δ_0=-0.05

Correlation Structure		**Type I error rate (%)**									**Power (%)**								
r_C	r_{CE}	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**
0	0	2.5	2.7^	2.9^	2.8^	2.9^	2.9^	2.9^	2.6	2.5	26.7	26.8	27.1	26.9	28.8	28.8	28.9	27.4	26.4
0.2	0	2.5	2.6	2.7^	2.6	2.7^	2.7^	2.7^	2.5	2.3*	27.1	26.6	26.8	26.5	28.3	28.3	28.3	26.8	25.9
0.4	0	2.7^	2.6	2.7^	2.7^	2.7^	2.7^	2.7^	2.4	2.3*	27.0	26.0	25.3	25.1	27.2	27.1	27.2	25.8	24.8
0.6	0	3.0^	2.6	2.8^	2.7^	2.8^	2.8^	2.8^	2.6	2.4	27.4	25.7	24.9	24.7	26.3	26.4	26.5	24.9	24.0
0.8	0	3.3^	2.6	2.7^	2.7^	2.7^	2.8^	[99]	2.5	2.4	27.4	24.8	24.0	23.8	25.1	25.1	[99]	23.9	22.8
0	0.2	1.8*	2.7^	2.8^	2.8^	2.8^	2.8^	3.0^	2.6	2.4	25.0	29.7	30.1	29.9	31.9	31.9	32.4	30.2	29.2
0.2	0.2	1.9*	2.6	2.8^	2.8^	2.8^	2.8^	2.9^	2.5	2.4	25.4	29.2	29.0	28.9	31.0	31.0	31.3	29.5	28.4
0.4	0.2	2.0*	2.6	2.7^	2.6	2.8^	2.8^	2.8^	2.6	2.4	25.8	28.7	28.3	28.2	29.9	30.0	30.2	28.5	27.4
0.6	0.2	2.2*	2.6	2.7^	2.7^	2.7^	2.7^	2.8^	2.5	2.3*	26.3	28.2	27.2	27.1	28.9	29.1	[99]	27.6	26.6
0.8	0.2	2.3*	2.4	2.5	2.5	2.5	2.5	[99]	2.3*	2.1*	27.0	28.2	26.5	26.4	28.3	28.5	[99]	26.8	25.6
0.4	0.4	1.3*	2.7^	2.8^	2.8^	2.8^	2.8^	2.9^	2.6	2.4	24.3	33.3	32.9	33.0	35.0	35.2	35.7	33.4	32.1
0.6	0.4	1.5*	2.6	2.7^	2.7^	2.8^	2.8^	2.8^	2.6	2.4	24.9	32.8	31.9	32.0	33.7	33.9	34.3	32.0	30.6
0	0.5	0.7*	2.7^	2.9^	2.9^	2.9^	2.8^	3.1^	2.6	2.3*	22.0	37.8	38.5	38.7	40.8	40.8	42.0	39.0	37.5
0.8	0.5	1.3*	2.6	2.6	2.7^	2.6	2.6	[99]	2.4	2.2*	24.1	34.0	32.3	32.7	34.2	34.9	[99]	32.7	31.2
0.6	0.6	0.7*	2.6	2.5	2.6	2.6	2.6	2.8^	2.5	2.2*	21.9	39.3	38.4	39.1	40.4	40.9	41.5	38.6	36.7
0.8	0.8	0.2*	2.8^	2.6	2.8^	2.9^	2.9^	[99]	2.7^	2.3*	18.1	51.9	49.2	51.2	51.3	53.5	[99]	49.4	46.7

Table 1.33. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k≤2, n_k~uniform, p_C=0.3, δ_0=-0.05

Correlation Structure		**Type I error rate (%)**									**Power (%)**								
r_C	r_{CE}	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**
0	0	2.6	2.7^	2.7^	2.6	2.7^	2.7^	2.8^	2.6	2.5	21.5	21.1	21.5	21.0	22.6	22.6	22.8	21.8	21.1
0.2	0	2.7^	2.5	2.7^	2.6	2.6	2.6	2.7^	2.5	2.4	22.3	21.3	21.2	20.8	22.5	22.6	22.7	21.7	21.0
0.4	0	2.8^	2.5	2.7^	2.6	2.6	2.7^	2.7^	2.5	2.4	22.5	20.8	20.6	20.2	21.6	21.6	21.8	20.9	20.2
0.6	0	3.0^	2.5	2.6	2.5	2.6	2.6	2.7^	2.5	2.3*	22.5	20.2	19.7	19.3	20.8	20.8	21.0	20.1	19.3
0.8	0	3.3^	2.4	2.5	2.5	2.6	2.6	2.6	2.5	2.3*	23.2	20.0	19.2	18.8	20.1	20.3	20.5	19.5	18.8
0	0.2	1.7*	2.6	2.7^	2.6	2.6	2.7^	2.8^	2.5	2.4	19.9	24.0	24.3	23.9	25.7	25.7	26.1	25.0	24.2
0.2	0.2	1.8*	2.5	2.7^	2.6	2.7^	2.7^	2.8^	2.6	2.4	20.3	23.5	23.8	23.4	25.1	25.0	25.3	24.2	23.4
0.4	0.2	2.0*	2.5	2.7^	2.6	2.7^	2.7^	2.8^	2.6	2.4	20.5	23.1	22.7	22.4	24.1	24.2	24.4	23.3	22.4
0.6	0.2	2.2*	2.5	2.6	2.5	2.7^	2.6	2.7^	2.5	2.4	21.8	23.3	22.8	22.4	23.8	24.0	24.2	23.1	22.2
0.8	0.2	2.5	2.5	2.6	2.5	2.6	2.6	[99]	2.5	2.3*	22.0	22.3	21.3	21.0	22.6	22.8	23.1	21.9	20.9
0.4	0.4	1.2*	2.6	2.7^	2.6	2.7^	2.7^	2.8^	2.6	2.4	18.8	27.1	27.0	26.8	28.3	28.5	28.9	27.6	26.4
0.6	0.4	1.4*	2.6	2.8^	2.7^	2.7^	2.8^	2.9^	2.6	2.4	19.3	26.3	26.0	25.8	27.1	27.4	27.7	26.4	25.2
0	0.5	0.6*	2.5	2.5	2.5	2.6	2.6	2.8^	2.5	2.3*	16.2	30.6	32.0	31.8	33.5	33.3	34.4	32.7	31.2
0.8	0.5	1.3*	2.6	2.6	2.6	2.7^	2.7^	2.8^	2.6	2.4	18.8	28.1	26.9	26.8	28.1	28.8	29.1	27.4	26.1
0.6	0.6	0.7*	2.6	2.6	2.6	2.7^	2.7^	2.9^	2.6	2.4	16.5	32.5	31.8	31.9	33.2	33.9	34.4	32.4	30.9
0.8	0.8	0.2*	2.7^	2.5	2.7^	2.7^	2.8^	3.0^	2.7^	2.4	12.4	43.4	41.9	43.0	42.8	45.1	45.8	42.0	39.7

Table 1.34. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤2, n_k~uniform, p_C=0.1, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.7^	2.8^	2.9^	2.8^	2.9^	2.9^	2.6	2.4	79.6	79.9	78.2	78.7	81.7	81.7	81.8	80.2	79.6
0.2	0	2.5	2.7^	2.8^	2.8^	2.8^	2.8^	2.8^	2.6	2.4	79.1	79.1	77.3	77.8	80.8	80.8	80.8	79.2	78.5
0.4	0	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.7^	2.4	2.3*	78.7	78.6	76.2	76.8	79.8	79.8	79.7	78.2	77.4
0.6	0	2.8^	2.6	2.7^	2.8^	2.7^	2.7^	2.7^	2.5	2.3*	78.7	78.3	75.3	75.9	79.0	79.0	79.0	77.2	76.5
0.8	0	3.0^	2.6	2.7^	2.7^	2.6	2.7^	2.6	2.4	2.2*	78.7	77.5	74.0	74.8	77.6	77.8	77.9	75.8	75.1
0	0.2	1.9*	2.7^	2.8^	2.9^	2.9^	2.9^	2.9^	2.6	2.5	80.6	83.3	81.8	82.4	85.1	85.1	85.3	83.7	83.2
0.2	0.2	1.9*	2.7^	2.7^	2.8^	2.8^	2.8^	2.8^	2.5	2.3*	80.4	83.0	81.4	82.0	84.5	84.5	84.7	83.1	82.4
0.4	0.2	2.1*	2.7^	2.7^	2.8^	2.8^	2.7^	2.8^	2.5	2.4	80.3	82.5	80.4	81.1	83.7	83.7	83.8	82.2	81.5
0.6	0.2	2.2*	2.6	2.7^	2.8^	2.7^	2.7^	2.7^	2.4	2.3*	79.7	81.6	79.0	79.7	82.3	82.4	82.4	80.7	80.0
0.8	0.2	2.5	2.7^	2.6	2.7^	2.7^	2.7^	2.8^	2.4	2.3*	79.3	80.9	77.4	78.2	80.9	81.2	81.3	79.2	78.5
0.4	0.4	1.4*	2.7^	2.7^	2.8^	2.8^	2.8^	2.9^	2.5	2.4	81.8	87.3	85.5	86.2	88.3	88.4	88.6	87.0	86.3
0.6	0.4	1.6*	2.6	2.6	2.8^	2.7^	2.7^	2.8^	2.4	2.3*	81.2	86.4	84.1	84.9	86.9	87.1	87.2	85.5	84.7
0	0.5	0.9*	2.7^	2.7^	2.9^	2.8^	2.8^	3.0^	2.5	2.3*	83.5	90.9	90.3	90.9	92.5	92.4	92.8	91.4	90.9
0.8	0.5	1.4*	2.7^	2.6	2.7^	2.7^	2.7^	2.8^	2.4	2.3*	81.8	88.4	85.4	86.4	88.1	88.7	88.8	86.8	86.1
0.6	0.6	0.9*	2.7^	2.7^	2.9^	2.8^	2.9^	3.0^	2.6	2.4	83.4	92.2	90.6	91.4	92.6	93.0	93.1	91.7	91.1
0.8	0.8	0.3*	3.1^	2.8^	3.2^	3.3^	3.2^	3.6^	2.9^	2.6	87.4	97.8	96.9	97.5	97.7	98.1	98.1	97.1	96.8

Table 1.35. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, $n_k \leq 2$, n_k~uniform, p_C=0.2, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.6	2.7^	2.6	2.6	2.6	2.7^	2.5	2.5	56.7	55.7	55.2	55.1	58.1	58.1	58.2	57.1	56.5
0.2	0	2.7^	2.7^	2.8^	2.7^	2.7^	2.7^	2.8^	2.6	2.5	57.1	55.6	54.4	54.3	57.4	57.4	57.4	56.5	55.8
0.4	0	2.8^	2.5	2.6	2.5	2.7^	2.6	2.6	2.5	2.4	56.4	54.6	53.0	52.9	55.7	55.7	55.9	54.7	54.0
0.6	0	3.0^	2.6	2.6	2.5	2.6	2.6	2.7^	2.5	2.4	56.3	53.6	51.5	51.4	54.3	54.3	54.4	53.4	52.7
0.8	0	3.2^	2.5	2.6	2.5	2.5	2.6	2.5	2.4	2.3*	56.4	53.0	50.2	50.1	52.6	52.9	53.2	51.7	50.8
0	0.2	1.8*	2.6	2.6	2.6	2.6	2.6	2.7^	2.5	2.4	57.6	61.5	61.2	61.2	64.1	64.0	64.3	63.3	62.5
0.2	0.2	2.0*	2.7^	2.7^	2.7^	2.7^	2.7^	2.8^	2.6	2.5	57.3	60.8	60.0	60.0	62.8	62.9	62.9	61.9	61.1
0.4	0.2	1.9*	2.4	2.5	2.5	2.5	2.5	2.5	2.4	2.3*	57.1	60.0	58.7	58.7	61.3	61.4	61.6	60.3	59.5
0.6	0.2	2.2*	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.3*	57.0	59.0	57.0	57.0	59.6	59.8	60.0	58.7	57.9
0.8	0.2	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.4	56.9	58.0	55.0	55.1	57.6	58.1	58.5	56.8	55.8
0.4	0.4	1.2*	2.5	2.6	2.6	2.6	2.5	2.6	2.4	2.3*	57.9	66.9	65.9	66.1	68.5	68.7	68.9	67.7	66.8
0.6	0.4	1.6*	2.6	2.8^	2.8^	2.8^	2.7^	2.8^	2.6	2.5	57.2	65.7	63.7	63.9	66.3	66.7	67.0	65.4	64.4
0	0.5	0.7*	2.6	2.7^	2.7^	2.7^	2.7^	2.9^	2.6	2.4	58.5	73.5	74.1	74.4	76.4	76.6	77.3	75.7	74.8
0.8	0.5	1.2*	2.6	2.6	2.6	2.6	2.6	2.7^	2.5	2.3*	57.5	68.6	65.9	66.2	68.0	69.1	69.4	67.0	66.0
0.6	0.6	0.7*	2.6	2.6	2.7^	2.7^	2.7^	2.7^	2.6	2.4	57.9	75.4	74.0	74.5	76.1	76.9	77.1	75.3	74.3
0.8	0.8	0.2*	2.7^	2.6	2.8^	2.8^	2.8^	3.0^	2.7^	2.4	59.8	88.8	87.3	88.0	88.0	89.7	89.9	87.4	86.5

Table 1.36. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤2, n_k~uniform, p_C=0.3, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.6	2.6	2.5	2.6	2.6	2.6	2.5	2.4	46.8	45.4	45.4	45.0	47.4	47.5	47.5	47.0	46.4
0.2	0	2.6	2.5	2.6	2.5	2.5	2.5	2.5	2.5	2.4	46.8	44.8	44.0	43.7	46.4	46.4	46.5	46.0	45.4
0.4	0	2.9^	2.5	2.6	2.6	2.6	2.6	2.6	2.5	2.5	46.4	43.6	42.5	42.2	44.7	44.8	44.9	44.2	43.6
0.6	0	3.0^	2.4	2.5	2.5	2.5	2.5	2.5	2.5	2.4	46.9	43.4	42.0	41.7	44.0	44.1	44.3	43.5	42.9
0.8	0	3.3^	2.5	2.5	2.5	2.5	2.5	2.5	2.4	2.4	46.4	42.0	39.9	39.7	41.8	42.0	42.3	41.4	40.7
0	0.2	1.7*	2.5	2.6	2.6	2.6	2.6	2.6	2.6	2.5	46.4	50.9	51.0	50.7	53.5	53.4	53.6	53.0	52.3
0.2	0.2	1.9*	2.6	2.7^	2.7^	2.7^	2.7^	2.7^	2.6	2.5	46.3	50.1	49.6	49.3	52.0	52.1	52.3	51.6	50.8
0.4	0.2	2.0*	2.5	2.6	2.5	2.6	2.5	2.6	2.5	2.4	46.4	49.3	48.2	47.9	50.2	50.4	50.6	49.7	49.0
0.6	0.2	2.0*	2.4	2.4	2.4	2.4	2.4	2.4	2.3*	2.2*	46.5	48.2	46.6	46.4	48.7	49.0	49.2	48.2	47.5
0.8	0.2	2.5	2.5	2.5	2.4	2.6	2.5	2.5	2.5	2.4	46.6	47.3	44.9	44.7	47.0	47.4	47.9	46.5	45.7
0.4	0.4	1.2*	2.5	2.4	2.4	2.5	2.5	2.5	2.4	2.3*	45.7	56.1	55.5	55.4	57.5	57.8	58.1	57.1	56.2
0.6	0.4	1.4*	2.6	2.6	2.6	2.6	2.6	2.7^	2.6	2.4	46.2	55.3	53.9	53.8	55.7	56.3	56.4	55.3	54.4
0	0.5	0.6*	2.6	2.5	2.5	2.6	2.6	2.7^	2.6	2.4	45.9	63.4	64.9	64.8	66.8	66.9	67.7	66.4	65.6
0.8	0.5	1.3*	2.6	2.5	2.5	2.6	2.6	2.7^	2.6	2.4	46.0	58.2	55.7	55.8	57.4	58.6	58.9	56.9	55.9
0.6	0.6	0.7*	2.6	2.5	2.5	2.5	2.6	2.7^	2.5	2.4	45.6	65.5	64.9	65.0	66.2	67.2	67.4	65.7	64.6
0.8	0.8	0.2*	2.6	2.5	2.6	2.6	2.6	2.7^	2.6	2.4	44.9	80.8	79.3	80.0	79.4	82.0	82.2	79.0	77.9

Table 1.37. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k≤2, n_k~uniform, p_C=0.1, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.4	2.6	2.8^	2.8^	2.7^	2.7^	2.7^	2.5	2.5	97.8	97.6	97.1	97.2	98.0	98.0	98.0	97.9	97.8
0.2	0	2.5	2.6	2.7^	2.8^	2.6	2.6	2.6	2.5	2.4	97.8	97.6	97.0	97.1	98.0	98.0	98.0	97.8	97.7
0.4	0	2.7^	2.7^	2.6	2.7^	2.7^	2.7^	2.7^	2.5	2.5	97.6	97.4	96.6	96.7	97.6	97.7	97.7	97.4	97.3
0.6	0	2.8^	2.6	2.6	2.6	2.7^	2.7^	2.6	2.5	2.4	97.4	97.1	96.3	96.4	97.3	97.3	97.3	97.1	97.0
0.8	0	3.0^	2.6	2.6	2.6	2.6	2.6	2.5	2.4	2.3*	97.2	96.9	95.6	95.7	96.8	96.9	97.0	96.5	96.4
0	0.2	1.9*	2.6	2.7^	2.8^	2.7^	2.7^	2.7^	2.5	2.4	98.2	98.5	98.2	98.2	98.8	98.8	98.8	98.7	98.6
0.2	0.2	2.0*	2.6	2.7^	2.8^	2.6	2.6	2.7^	2.5	2.4	98.2	98.5	98.1	98.3	98.8	98.8	98.8	98.6	98.5
0.4	0.2	2.2*	2.7^	2.7^	2.8^	2.8^	2.8^	2.8^	2.6	2.5	98.1	98.3	97.8	97.9	98.6	98.6	98.6	98.4	98.3
0.6	0.2	2.4	2.7^	2.8^	2.8^	2.8^	2.8^	2.8^	2.6	2.5	97.8	98.0	97.4	97.5	98.2	98.2	98.2	98.0	97.9
0.8	0.2	2.5	2.6	2.6	2.7^	2.6	2.6	2.7^	2.5	2.4	97.7	98.0	97.0	97.1	97.9	97.9	98.0	97.6	97.5
0.4	0.4	1.4*	2.8^	2.7^	2.8^	2.7^	2.7^	2.8^	2.5	2.4	98.6	99.2	99.0	99.0	99.3	99.3	99.3	99.2	99.1
0.6	0.4	1.6*	2.6	2.6	2.7^	2.7^	2.6	2.7^	2.5	2.4	98.5	99.1	98.7	98.7	99.1	99.2	99.2	99.0	99.0
0	0.5	0.9*	2.7^	2.7^	2.8^	2.8^	2.8^	2.9^	2.6	2.5	99.1	99.6	99.5	99.6	99.8	99.7	99.8	99.7	99.7
0.8	0.5	1.4*	2.8^	2.6	2.7^	2.8^	2.8^	2.8^	2.6	2.4	98.6	99.3	98.9	99.0	99.2	99.3	99.3	99.1	99.1
0.6	0.6	0.8*	2.5	2.5	2.6	2.6	2.6	2.7^	2.4	2.3*	99.1	99.8	99.6	99.7	99.8	99.8	99.8	99.7	99.7
0.8	0.8	0.3*	2.8^	2.6	2.9^	2.8^	2.8^	3.1^	2.6	2.4	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.38. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k≤2, n_k~uniform, p_C=0.2, δ_0=-0.05

Correlation Structure		**Type I error rate (%)**									**Power (%)**								
r_C	r_{CE}	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**
0	0	2.4	2.5	2.6	2.6	2.6	2.6	2.6	2.5	2.4	85.6	84.6	83.8	83.8	86.2	86.2	86.1	85.8	85.6
0.2	0	2.6	2.5	2.6	2.6	2.6	2.6	2.6	2.5	2.5	85.3	84.2	83.1	83.0	85.3	85.3	85.3	85.0	84.7
0.4	0	2.7^	2.5	2.5	2.5	2.5	2.5	2.5	2.4	2.4	85.1	83.6	81.9	81.9	84.4	84.5	84.5	84.1	83.8
0.6	0	3.0^	2.6	2.6	2.6	2.5	2.5	2.5	2.5	2.4	84.7	82.8	80.6	80.6	83.1	83.3	83.4	82.8	82.4
0.8	0	3.1^	2.4	2.5	2.5	2.4	2.4	2.4	2.4	2.3*	84.1	82.0	79.3	79.3	81.5	81.8	82.2	81.1	80.7
0	0.2	1.7*	2.5	2.6	2.6	2.6	2.6	2.6	2.5	2.4	87.3	88.8	88.5	88.6	90.3	90.4	90.5	90.1	89.9
0.2	0.2	1.9*	2.6	2.7^	2.7^	2.6	2.6	2.6	2.5	2.4	86.7	88.1	87.4	87.4	89.4	89.5	89.5	89.1	88.9
0.4	0.2	2.0*	2.5	2.5	2.5	2.5	2.5	2.5	2.4	2.3*	86.3	87.4	86.4	86.4	88.3	88.3	88.4	88.0	87.7
0.6	0.2	2.2*	2.5	2.5	2.5	2.5	2.5	2.5	2.4	2.3*	86.2	87.0	85.4	85.4	87.4	87.6	87.7	87.0	86.7
0.8	0.2	2.5	2.5	2.6	2.6	2.5	2.5	2.5	2.5	2.4	85.7	86.4	83.9	84.0	86.0	86.4	86.6	85.7	85.3
0.4	0.4	1.3*	2.5	2.6	2.7^	2.6	2.6	2.6	2.5	2.4	88.4	92.2	91.6	91.6	92.9	93.0	93.1	92.7	92.4
0.6	0.4	1.4*	2.6	2.6	2.6	2.5	2.5	2.6	2.4	2.3*	87.8	91.6	90.4	90.5	91.8	92.1	92.2	91.5	91.3
0	0.5	0.7*	2.6	2.7^	2.7^	2.6	2.6	2.7^	2.6	2.5	90.9	95.6	95.9	95.9	96.6	96.6	96.8	96.5	96.3
0.8	0.5	1.4*	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.4	88.3	93.0	91.5	91.6	92.8	93.3	93.4	92.5	92.2
0.6	0.6	0.8*	2.7^	2.6	2.6	2.6	2.7^	2.8^	2.6	2.4	90.9	96.4	96.0	96.1	96.5	96.9	96.9	96.4	96.2
0.8	0.8	0.3*	2.8^	2.6	2.7^	2.7^	2.8^	2.9^	2.7^	2.5	93.6	99.4	99.2	99.2	99.2	99.5	99.5	99.2	99.1

Table 1.39. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k≤2, n_k~uniform, p_C=0.3, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.4	2.4	2.5	2.5	2.5	2.5	2.5	2.5	2.4	75.6	74.0	73.4	73.2	76.0	76.0	76.0	75.8	75.5
0.2	0	2.7^	2.5	2.6	2.5	2.6	2.6	2.6	2.6	2.5	75.1	72.9	71.8	71.7	74.5	74.5	74.6	74.3	73.9
0.4	0	2.8^	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.4	75.3	72.7	71.1	71.0	73.8	73.9	73.9	73.6	73.2
0.6	0	3.1^	2.6	2.6	2.5	2.6	2.5	2.6	2.5	2.5	74.6	71.9	69.4	69.3	71.8	72.0	72.3	71.6	71.2
0.8	0	3.4^	2.6	2.5	2.5	2.5	2.6	2.6	2.5	2.4	74.6	71.0	67.8	67.7	70.4	70.8	71.3	70.2	69.8
0	0.2	1.7*	2.4	2.5	2.5	2.6	2.5	2.5	2.5	2.4	77.1	79.7	79.7	79.6	81.8	81.8	82.0	81.7	81.3
0.2	0.2	1.9*	2.5	2.6	2.6	2.6	2.6	2.6	2.5	2.5	77.2	79.4	78.8	78.7	81.0	81.1	81.1	80.8	80.5
0.4	0.2	2.1*	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	76.6	78.3	77.1	77.0	79.4	79.5	79.6	79.2	78.9
0.6	0.2	2.4	2.7^	2.6	2.5	2.7^	2.7^	2.7^	2.6	2.5	76.0	77.3	75.4	75.3	77.6	78.0	78.1	77.4	77.1
0.8	0.2	2.5	2.5	2.5	2.5	2.6	2.6	2.6	2.5	2.5	75.3	76.0	73.2	73.1	75.5	76.0	76.4	75.2	74.8
0.4	0.4	1.2*	2.4	2.6	2.6	2.5	2.5	2.5	2.5	2.4	78.4	85.0	84.2	84.2	85.9	86.3	86.3	85.7	85.3
0.6	0.4	1.4*	2.5	2.5	2.5	2.5	2.4	2.5	2.5	2.4	77.9	84.1	82.6	82.6	84.5	85.0	85.0	84.3	83.9
0	0.5	0.6*	2.5	2.6	2.6	2.6	2.6	2.6	2.6	2.5	81.0	90.2	91.2	91.2	92.2	92.2	92.6	92.1	91.8
0.8	0.5	1.3*	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	78.6	86.6	84.7	84.7	85.9	86.9	87.1	85.7	85.3
0.6	0.6	0.6*	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	81.4	91.9	91.3	91.4	92.1	92.6	92.7	92.0	91.6
0.8	0.8	0.2*	2.7^	2.6	2.7^	2.7^	2.7^	2.8^	2.7^	2.6	84.8	97.9	97.7	97.8	97.5	98.2	98.2	97.5	97.3

Table 1.40. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, $n_k \le 2$, $n_k \sim$uniform, $p_C=0.05$, $\delta_0=-0.02$

Correlation Structure		**Type I error rate (%)**									**Power (%)**									
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.3*	2.6	2.8^	2.9^	2.7^	2.7^	2.8^	2.4	2.3*	58.2	59.6	57.9	58.8	61.6	61.6	61.7	59.2	58.3
0.2	0	2.5	2.8^	2.9^	3.1^	2.9^	2.9^	2.9^	2.6	2.5	58.0	59.0	57.0	57.9	60.9	60.9	60.8	58.4	57.6
0.4	0	2.6	2.7^	2.8^	2.9^	2.8^	2.8^	2.8^	2.5	2.4	57.7	58.6	56.0	57.1	59.7	59.7	59.7	57.3	56.5
0.6	0	2.7^	2.7^	2.7^	2.8^	2.7^	2.7^	2.7^	2.4	2.3*	57.5	57.6	54.9	55.8	58.4	58.3	58.3	56.0	55.0
0.8	0	2.9^	2.6	2.7^	2.8^	2.6	2.6	2.6	2.3*	2.2*	57.8	56.8	53.2	54.2	56.9	57.0	57.1	54.6	53.5
0	0.2	2.1*	2.8^	2.9^	3.0^	3.0^	3.0^	3.0^	2.6	2.5	58.3	62.3	60.7	61.7	64.6	64.6	64.7	62.1	61.2
0.2	0.2	2.0*	2.7^	2.8^	3.0^	2.8^	2.8^	2.9^	2.5	2.3*	58.2	61.9	60.0	61.1	63.7	63.7	63.8	61.3	60.5
0.4	0.2	2.1*	2.7^	2.8^	2.9^	2.8^	2.8^	2.8^	2.5	2.4	58.3	61.6	58.9	60.0	62.9	62.8	62.9	60.5	59.5
0.6	0.2	2.4	2.8^	2.8^	2.9^	2.9^	2.9^	2.9^	2.5	2.4	57.8	60.6	57.6	58.7	61.3	61.4	61.4	58.8	57.8
0.8	0.2	2.4	2.7^	2.6	2.7^	2.6	2.6	2.6	2.3*	2.2*	57.7	59.4	55.5	56.7	59.2	59.5	59.6	56.9	55.9
0.4	0.4	1.7*	2.8^	2.8^	3.0^	2.9^	2.9^	3.0^	2.5	2.3*	58.1	65.6	63.5	64.7	67.3	67.4	67.6	64.8	63.8
0.6	0.4	1.8*	2.8^	2.8^	3.0^	2.8^	2.8^	2.8^	2.5	2.4	58.4	65.0	62.1	63.4	65.9	66.1	66.3	63.4	62.4
0	0.5	1.3*	2.8^	2.9^	3.1^	2.9^	2.9^	3.1^	2.6	2.4	58.9	70.4	69.4	70.6	73.2	73.1	73.7	70.8	69.7
0.8	0.5	1.6*	2.6	2.6	2.8^	2.7^	2.7^	2.7^	2.4	2.2*	58.6	66.7	63.3	64.8	67.0	67.5	67.6	64.5	63.4
0.6	0.6	1.0*	2.8^	2.7^	2.9^	2.8^	2.8^	3.0^	2.5	2.3*	58.5	72.6	70.1	71.5	73.5	73.8	74.2	71.1	69.9
0.8	0.8	0.4*	3.0^	2.7^	3.1^	3.0^	3.1^	3.2^	2.7^	2.4	60.3	84.9	82.0	83.8	84.5	85.8	86.0	82.3	81.1

Table 1.41. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤2, n_k~uniform, p_C=0.05, δ_0=-0.02

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.9^	3.0^	3.2^	3.1^	3.1^	3.1^	2.6	2.4	32.5	35.3	34.2	35.5	37.0	36.9	37.0	33.7	32.7
0.2	0	2.3*	2.7^	2.9^	3.1^	2.9^	2.9^	2.9^	2.5	2.3*	32.3	34.7	33.5	34.8	36.4	36.5	36.5	33.1	32.0
0.4	0	2.5	3.0^	2.9^	3.1^	3.0^	3.0^	3.0^	2.5	2.3*	32.4	34.8	32.8	34.0	35.8	35.8	35.8	32.5	31.3
0.6	0	2.6	2.8^	2.9^	3.0^	2.9^	2.8^	2.9^	2.4	2.3*	32.6	33.8	31.8	33.0	34.6	34.5	34.5	31.3	30.1
0.8	0	3.0^	2.9^	2.9^	3.1^	2.9^	2.9^	[99]	2.5	2.3*	32.7	32.8	30.1	31.4	32.9	33.0	[99]	29.8	28.6
0	0.2	2.0*	2.9^	3.1^	3.3^	3.1^	3.1^	3.2^	2.6	2.4	31.1	36.7	35.5	36.9	38.7	38.7	39.0	35.0	33.9
0.2	0.2	2.1*	2.9^	3.0^	3.2^	3.1^	3.1^	3.1^	2.6	2.4	31.8	36.7	35.4	36.7	38.6	38.6	38.8	35.1	33.9
0.4	0.2	2.1*	2.9^	2.9^	3.2^	3.0^	3.0^	3.1^	2.5	2.4	32.0	36.5	34.8	36.2	37.8	37.8	38.0	34.4	33.1
0.6	0.2	2.2*	2.7^	2.6	2.8^	2.8^	2.8^	2.8^	2.4	2.2*	32.2	35.9	33.5	35.0	36.7	36.7	36.8	33.2	32.0
0.8	0.2	2.4	2.7^	2.6	2.8^	2.7^	2.7^	[99]	2.2*	2.0*	32.3	34.6	32.1	33.6	34.8	35.0	[99]	31.6	30.4
0.4	0.4	1.6*	3.0^	3.0^	3.2^	3.2^	3.1^	3.2^	2.6	2.4	30.7	39.6	37.8	39.5	41.2	41.2	41.5	37.4	35.9
0.6	0.4	1.7*	2.7^	2.8^	3.0^	2.9^	2.9^	3.0^	2.4	2.2*	31.4	39.2	36.9	38.7	40.1	40.2	40.4	36.5	35.0
0	0.5	1.1*	2.8^	2.9^	3.2^	3.1^	3.0^	3.2^	2.5	2.3*	30.1	43.6	42.7	44.6	46.2	46.2	47.1	42.2	40.6
0.8	0.5	1.6*	2.8^	2.6	3.0^	2.9^	2.9^	[99]	2.4	2.2*	30.8	40.1	36.9	38.9	40.1	40.5	[99]	36.5	34.8
0.6	0.6	1.1*	3.0^	2.9^	3.3^	3.1^	3.1^	3.3^	2.6	2.3*	29.8	45.4	42.8	45.3	46.5	46.7	47.2	42.5	40.7
0.8	0.8	0.4*	3.1^	2.8^	3.4^	3.2^	3.1^	[99]	2.6	2.2*	26.6	57.0	52.7	56.7	57.2	58.3	[99]	52.6	49.8

Table 1.42. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, $n_k \le 2$, n_k~uniform, p_C=0.2, δ_0=-0.02

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.4	2.6	2.6	2.6	2.6	2.6	2.7^	2.5	2.4	13.3	13.3	13.5	13.5	14.2	14.2	14.2	13.7	13.3
0.2	0	2.6	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.5	2.4	13.4	13.1	13.2	13.2	13.6	13.6	13.7	13.1	12.7
0.4	0	2.8^	2.6	2.7^	2.7^	2.6	2.6	2.6	2.5	2.4	14.1	13.2	13.0	13.0	13.5	13.5	13.6	13.0	12.7
0.6	0	2.9^	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.4	14.3	12.8	12.4	12.4	13.0	13.0	13.0	12.6	12.2
0.8	0	3.3^	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.4	14.5	12.4	11.8	11.8	12.3	12.4	12.5	11.9	11.5
0	0.2	1.8*	2.6	2.7^	2.7^	2.7^	2.7^	2.8^	2.6	2.5	11.9	14.8	15.1	15.0	15.7	15.7	15.8	15.2	14.7
0.2	0.2	1.9*	2.5	2.7^	2.7^	2.6	2.5	2.6	2.4	2.3*	12.4	14.6	14.7	14.6	15.3	15.3	15.4	14.8	14.3
0.4	0.2	2.1*	2.5	2.5	2.5	2.6	2.6	2.6	2.5	2.4	12.5	14.2	14.2	14.1	14.8	14.7	14.9	14.2	13.8
0.6	0.2	2.2*	2.5	2.6	2.6	2.5	2.6	2.6	2.4	2.3*	12.7	13.7	13.4	13.4	14.0	14.0	14.1	13.6	13.1
0.8	0.2	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.4	13.5	13.9	13.2	13.2	13.7	13.9	14.0	13.3	12.8
0.4	0.4	1.3*	2.6	2.6	2.6	2.7^	2.6	2.7^	2.6	2.4	10.6	16.2	16.1	16.1	16.8	16.8	17.0	16.2	15.6
0.6	0.4	1.5*	2.6	2.7^	2.7^	2.7^	2.6	2.6	2.5	2.4	11.0	15.7	15.4	15.4	15.9	16.1	16.2	15.4	14.8
0	0.5	0.7*	2.6	2.8^	2.8^	2.7^	2.7^	2.9^	2.6	2.5	8.7	18.1	18.7	18.7	19.5	19.4	20.0	18.9	18.2
0.8	0.5	1.3*	2.4	2.4	2.4	2.5	2.4	2.5	2.4	2.2*	10.8	16.6	16.0	16.0	16.5	16.8	16.9	15.9	15.3
0.6	0.6	0.8*	2.7^	2.7^	2.7^	2.7^	2.7^	2.8^	2.6	2.4	8.4	18.6	18.4	18.5	18.8	19.2	19.6	18.1	17.4
0.8	0.8	0.2*	2.5	2.5	2.6	2.7^	2.6	2.7^	2.6	2.3*	5.3	24.7	24.1	24.5	24.3	25.4	25.9	23.6	22.3

Table 1.43. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤2, n_k~uniform, p_C=0.2, δ_0=-0.1

Correlation Structure		***Type I error rate (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.5	2.6	2.5	2.6	2.6	2.6	2.5	2.4
0.2	0	2.5	2.5	2.6	2.6	2.5	2.5	2.6	2.5	2.3*
0.4	0	2.8^	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.5
0.6	0	3.0^	2.5	2.6	2.5	2.6	2.6	2.6	2.5	2.4
0.8	0	3.1^	2.4	2.5	2.5	2.5	2.5	2.4	2.4	2.3*
0	0.2	1.7*	2.5	2.7^	2.6	2.6	2.6	2.6	2.5	2.4
0.2	0.2	1.8*	2.5	2.6	2.6	2.6	2.6	2.6	2.5	2.4
0.4	0.2	2.0*	2.5	2.5	2.5	2.6	2.6	2.6	2.5	2.4
0.6	0.2	2.3*	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.6	2.5
0.8	0.2	2.5	2.5	2.6	2.6	2.6	2.6	2.6	2.6	2.5
0.4	0.4	1.2*	2.5	2.7^	2.7^	2.5	2.6	2.6	2.5	2.4
0.6	0.4	1.5*	2.6	2.6	2.6	2.6	2.7^	2.7^	2.6	2.5
0	0.5	0.7*	2.5	2.5	2.6	2.6	2.6	2.7^	2.6	2.4
0.8	0.5	1.3*	2.6	2.6	2.7^	2.6	2.7^	2.7^	2.6	2.4
0.6	0.6	0.7*	2.6	2.5	2.6	2.7^	2.6	2.8^	2.6	2.4
0.8	0.8	0.3*	2.9^	2.6	2.9^	2.9^	3.0^	3.3^	2.9^	2.6

Power (%)

FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
99.0	98.8	98.5	98.5	99.0	99.0	99.0	99.0	98.9
98.8	98.6	98.3	98.3	98.8	98.8	98.8	98.8	98.7
98.7	98.5	98.1	98.1	98.7	98.7	98.7	98.6	98.5
98.6	98.3	97.7	97.7	98.4	98.4	98.4	98.3	98.2
98.5	98.2	97.4	97.4	98.2	98.3	98.3	98.1	98.0
99.3	99.4	99.2	99.2	99.6	99.6	99.6	99.5	99.5
99.2	99.3	99.1	99.1	99.4	99.4	99.4	99.4	99.3
99.0	99.2	99.0	99.0	99.3	99.3	99.3	99.2	99.2
99.1	99.2	98.8	98.8	99.3	99.3	99.3	99.2	99.2
98.9	99.0	98.5	98.5	99.0	99.0	99.0	98.9	98.9
99.6	99.8	99.7	99.7	99.8	99.8	99.8	99.8	99.8
99.4	99.7	99.5	99.5	99.7	99.7	99.7	99.7	99.7
99.8	99.9	99.9	99.9	100.0	99.9	100.0	100.0	99.9
99.6	99.8	99.7	99.8	99.8	99.8	99.9	99.8	99.8
99.8	100.0	99.9	99.9	99.9	100.0	100.0	99.9	99.9
100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.44. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, $n_k \le 2$, n_k~uniform, p_C=0.3, δ_0=-0.1

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.5	2.7^	2.6	2.6	2.6	2.6	2.5	2.5	96.3	95.6	95.3	95.2	96.4	96.4	96.4	96.4	96.2
0.2	0	2.6	2.4	2.5	2.4	2.4	2.4	2.4	2.4	2.3*	96.0	95.2	94.8	94.7	95.9	95.9	95.9	95.8	95.7
0.4	0	2.8^	2.5	2.6	2.5	2.5	2.6	2.6	2.5	2.4	95.8	95.1	94.3	94.2	95.5	95.5	95.5	95.4	95.3
0.6	0	3.0^	2.5	2.5	2.4	2.5	2.5	2.5	2.5	2.4	95.6	94.7	93.6	93.5	94.8	94.8	94.9	94.7	94.5
0.8	0	3.2^	2.5	2.5	2.4	2.5	2.5	2.5	2.5	2.4	95.2	94.1	92.6	92.5	93.9	94.0	94.3	93.8	93.6
0	0.2	1.7*	2.5	2.6	2.5	2.6	2.6	2.7^	2.6	2.5	97.3	97.6	97.6	97.5	98.1	98.1	98.2	98.1	98.0
0.2	0.2	1.8*	2.5	2.6	2.5	2.5	2.5	2.5	2.5	2.4	97.1	97.3	97.1	97.1	97.8	97.9	97.9	97.8	97.7
0.4	0.2	2.0*	2.5	2.6	2.5	2.5	2.5	2.6	2.5	2.4	96.7	97.1	96.6	96.6	97.4	97.4	97.5	97.4	97.2
0.6	0.2	2.3*	2.6	2.6	2.5	2.6	2.6	2.7^	2.6	2.5	96.6	96.8	96.0	96.0	97.0	97.1	97.1	96.9	96.8
0.8	0.2	2.5	2.5	2.6	2.5	2.6	2.6	2.6	2.6	2.4	96.3	96.6	95.5	95.4	96.4	96.6	96.7	96.3	96.2
0.4	0.4	1.2*	2.6	2.6	2.6	2.5	2.6	2.6	2.5	2.4	97.9	98.8	98.7	98.7	99.0	99.0	99.1	99.0	98.9
0.6	0.4	1.3*	2.5	2.6	2.6	2.6	2.6	2.6	2.6	2.4	97.8	98.7	98.4	98.4	98.8	98.9	98.9	98.8	98.7
0	0.5	0.7*	2.6	2.6	2.6	2.6	2.6	2.7^	2.6	2.5	98.9	99.6	99.7	99.7	99.8	99.8	99.8	99.8	99.8
0.8	0.5	1.2*	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.8^	2.6	98.0	99.1	98.7	98.8	99.0	99.2	99.2	99.0	99.0
0.6	0.6	0.6*	2.5	2.6	2.6	2.6	2.6	2.7^	2.6	2.5	98.7	99.7	99.6	99.6	99.7	99.7	99.8	99.7	99.7
0.8	0.8	0.2*	2.6	2.5	2.7^	2.7^	2.7^	2.8^	2.7^	2.5	99.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.45. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤2, n_k~uniform, p_C=0.5, δ_0=-0.1

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.6	2.5	2.6	2.5	2.6	2.6	2.6	2.6	2.5	93.6	92.2	92.0	91.8	93.5	93.5	93.5	93.6	93.4
0.2	0	2.7^	2.3*	2.4	2.3*	2.4	2.4	2.5	2.5	2.4	93.2	91.7	91.2	91.0	92.7	92.7	92.7	92.7	92.6
0.4	0	3.1^	2.5	2.6	2.4	2.6	2.6	2.6	2.6	2.5	92.9	91.3	90.3	90.1	91.9	91.9	92.0	91.9	91.7
0.6	0	3.2^	2.5	2.5	2.4	2.5	2.5	2.5	2.6	2.5	92.6	90.7	89.4	89.1	91.0	91.1	91.3	91.1	90.8
0.8	0	3.6^	2.5	2.5	2.4	2.5	2.5	2.5	2.6	2.5	92.0	90.0	88.0	87.7	89.6	89.8	90.2	89.7	89.4
0	0.2	1.6*	2.4	2.5	2.4	2.5	2.5	2.5	2.5	2.4	94.9	95.4	95.5	95.4	96.3	96.3	96.4	96.4	96.2
0.2	0.2	1.7*	2.4	2.3*	2.2*	2.3*	2.3*	2.4	2.4	2.3*	94.8	95.2	95.0	94.9	95.9	95.9	96.0	96.0	95.8
0.4	0.2	2.1*	2.5	2.5	2.4	2.5	2.5	2.6	2.6	2.5	94.4	94.8	94.3	94.2	95.3	95.4	95.4	95.3	95.2
0.6	0.2	2.4	2.5	2.5	2.4	2.6	2.6	2.6	2.6	2.5	94.0	94.2	93.4	93.2	94.4	94.6	94.7	94.5	94.3
0.8	0.2	2.7^	2.5	2.5	2.5	2.6	2.6	2.6	2.6	2.5	93.6	93.7	92.4	92.2	93.4	93.8	94.0	93.5	93.3
0.4	0.4	1.2*	2.5	2.4	2.4	2.5	2.5	2.5	2.6	2.4	96.0	97.6	97.4	97.4	97.8	97.9	97.9	97.9	97.8
0.6	0.4	1.4*	2.5	2.5	2.5	2.5	2.6	2.6	2.6	2.5	95.6	97.4	96.8	96.8	97.3	97.6	97.6	97.4	97.2
0	0.5	0.6*	2.5	2.4	2.4	2.4	2.5	2.5	2.5	2.4	97.4	99.0	99.2	99.2	99.4	99.4	99.5	99.4	99.4
0.8	0.5	1.3*	2.4	2.5	2.5	2.5	2.5	2.5	2.6	2.4	95.8	98.0	97.5	97.5	97.8	98.1	98.2	97.8	97.7
0.6	0.6	0.7*	2.5	2.5	2.6	2.6	2.6	2.6	2.6	2.5	97.3	99.3	99.2	99.2	99.3	99.4	99.4	99.3	99.3
0.8	0.8	0.2*	2.6	2.4	2.6	2.7^	2.6	2.8^	2.8^	2.6	98.9	100.0	99.9	99.9	99.9	100.0	100.0	99.9	99.9

Table 1.46. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, $n_k \leq 2$, $n_k \sim beta(2,3)$, $p_C=0.1$, $\delta_0=-0.05$

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.9^	3.0^	3.0^	3.0^	3.0^	3.1^	2.6	2.4	39.8	41.6	40.4	41.4	44.0	44.0	44.2	40.9	39.4
0.2	0	2.4	2.7^	2.9^	3.0^	2.9^	2.9^	[99]	2.5	2.3*	39.4	41.0	39.2	40.3	43.0	42.9	[99]	40.0	38.6
0.4	0	2.6	2.7^	2.8^	2.9^	2.9^	2.9^	[99]	2.5	2.3*	39.4	40.7	38.3	39.5	42.3	42.3	[99]	39.1	37.8
0.6	0	2.7^	2.7^	2.8^	2.8^	2.8^	2.8^	[99]	2.5	2.3*	40.1	40.6	38.2	39.3	41.7	41.6	[99]	38.7	37.2
0.8	0	2.8^	2.6	2.6	2.7^	2.6	2.6	[98]	2.3*	2.1*	40.2	40.1	36.7	37.9	40.4	40.5	[98]	37.4	35.8
0	0.2	1.9*	2.7^	2.8^	3.0^	2.9^	2.9^	3.0^	2.5	2.3*	39.5	44.9	43.3	44.7	47.6	47.5	47.8	44.2	42.6
0.2	0.2	1.8*	2.6	2.7^	2.9^	2.8^	2.8^	2.9^	2.4	2.2*	39.2	44.3	42.2	43.7	46.5	46.5	[99]	43.3	41.7
0.4	0.2	2.0*	2.8^	2.8^	3.0^	2.9^	2.9^	[99]	2.5	2.3*	39.2	43.8	41.6	43.1	45.7	45.8	[99]	42.5	40.8
0.6	0.2	2.1*	2.6	2.7^	2.8^	2.8^	2.8^	[99]	2.4	2.2*	39.0	43.3	40.5	42.0	44.3	44.4	[99]	41.1	39.4
0.8	0.2	2.4	2.8^	2.8^	2.9^	2.8^	2.8^	[98]	2.5	2.3*	39.3	42.7	39.0	40.6	43.0	43.1	[98]	39.8	37.9
0.4	0.4	1.5*	2.8^	2.8^	3.1^	3.0^	3.0^	[99]	2.6	2.3*	38.6	48.8	46.2	48.3	50.6	50.8	[99]	47.3	45.4
0.6	0.4	1.5*	2.6	2.6	2.8^	2.9^	2.8^	[99]	2.6	2.3*	38.4	48.2	44.8	46.8	49.1	49.3	[99]	45.8	43.8
0	0.5	1.0*	2.9^	2.9^	3.1^	3.1^	3.1^	3.4^	2.6	2.3*	38.0	52.9	51.3	53.6	56.0	56.0	[99]	52.6	50.6
0.8	0.5	1.3*	2.7^	2.6	2.9^	2.9^	2.8^	[98]	2.5	2.2*	37.9	50.5	45.9	48.4	50.5	51.2	[98]	47.1	44.8
0.6	0.6	0.8*	2.8^	2.6	3.0^	2.9^	2.9^	[99]	2.5	2.2*	37.5	56.2	52.3	55.4	57.3	57.7	[99]	53.7	51.4
0.8	0.8	0.3*	3.4^	2.8^	3.7^	3.3^	3.5^	[99]	3.0^	2.5	36.1	70.6	64.3	69.4	70.2	71.9	[99]	66.2	63.3

Table 1.47. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, $n_k \le 2$, $n_k \sim beta(2,3)$, $p_C=0.2$, $\delta_0=-0.05$

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.5	2.6	2.6	2.6	2.6	2.7^	2.4	2.3*	25.8	25.6	25.6	25.3	27.4	27.4	27.7	26.3	25.4
0.2	0	2.6	2.6	2.7^	2.7^	2.8^	2.8^	2.8^	2.6	2.4	26.1	25.8	25.6	25.4	27.1	27.1	27.2	26.1	25.2
0.4	0	2.7^	2.5	2.7^	2.7^	2.6	2.6	2.6	2.5	2.3*	26.0	25.2	24.5	24.3	26.1	26.1	[99]	25.1	24.1
0.6	0	2.9^	2.6	2.8^	2.7^	2.7^	2.7^	[99]	2.6	2.4	26.4	25.0	24.0	23.8	25.5	25.5	[99]	24.5	23.5
0.8	0	3.1^	2.5	2.6	2.5	2.6	2.7^	[99]	2.5	2.4	26.7	24.7	23.3	23.2	24.7	24.8	[99]	23.8	22.9
0	0.2	1.8*	2.6	2.7^	2.7^	2.7^	2.7^	2.9^	2.6	2.4	24.4	28.1	28.6	28.5	30.2	30.2	30.6	29.1	28.1
0.2	0.2	1.8*	2.5	2.5	2.5	2.6	2.6	2.7^	2.5	2.3*	24.6	28.1	27.9	27.8	29.6	29.6	29.8	28.5	27.4
0.4	0.2	2.0*	2.5	2.6	2.6	2.7^	2.7^	2.7^	2.6	2.4	24.8	27.7	27.1	27.1	28.9	29.0	29.0	27.8	26.6
0.6	0.2	2.2*	2.5	2.6	2.6	2.6	2.6	[99]	2.5	2.3*	25.3	27.5	26.6	26.6	28.2	28.3	[99]	27.2	26.1
0.8	0.2	2.2*	2.5	2.6	2.5	2.4	2.5	[99]	2.3*	2.1*	25.8	27.2	25.8	25.8	27.3	27.6	[99]	26.3	25.1
0.4	0.4	1.3*	2.5	2.5	2.5	2.7^	2.7^	[99]	2.6	2.4	22.9	31.4	31.1	31.2	32.7	33.0	33.1	31.6	30.2
0.6	0.4	1.3*	2.5	2.5	2.5	2.6	2.5	[99]	2.5	2.2*	23.5	31.0	30.1	30.2	31.9	32.2	[99]	30.8	29.4
0	0.5	0.7*	2.5	2.6	2.6	2.8^	2.7^	2.9^	2.6	2.4	21.2	35.1	35.7	36.0	37.9	37.8	38.7	36.6	35.1
0.8	0.5	1.2*	2.6	2.6	2.7^	2.7^	2.7^	[99]	2.6	2.4	22.4	32.7	30.8	31.3	32.6	33.3	[99]	31.5	30.0
0.6	0.6	0.8*	2.7^	2.7^	2.8^	2.8^	2.8^	[99]	2.6	2.3*	20.8	37.5	36.3	37.0	38.3	38.9	[99]	37.1	35.2
0.8	0.8	0.2*	2.9^	2.7^	3.0^	3.0^	3.0^	[99]	3.0^	2.5	17.3	50.1	47.0	49.1	49.1	51.5	[99]	47.7	45.1

Table 1.48. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k≤2, n_k~beta(2,3), p_C=0.3, δ_0=-0.05

Correlation Structure		**Type I error rate (%)**									**Power (%)**								
r_C	r_{CE}	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**
0	0	2.5	2.5	2.6	2.5	2.6	2.6	2.7^	2.5	2.4	20.9	20.1	20.4	20.0	21.6	21.6	21.7	21.1	20.4
0.2	0	2.7^	2.5	2.7^	2.6	2.7^	2.7^	2.7^	2.6	2.5	21.4	20.2	20.2	19.8	21.5	21.5	21.6	20.9	20.2
0.4	0	2.8^	2.5	2.6	2.5	2.6	2.6	2.7^	2.5	2.4	21.7	20.2	20.0	19.6	21.0	21.0	21.1	20.5	19.8
0.6	0	2.9^	2.4	2.6	2.5	2.5	2.6	2.6	2.4	2.3*	22.3	20.3	19.6	19.2	20.8	20.8	20.9	20.3	19.5
0.8	0	3.0^	2.3*	2.4	2.3*	2.4	2.4	[99]	2.4	2.2*	22.0	19.4	18.6	18.3	19.5	19.6	[99]	19.1	18.3
0	0.2	1.7*	2.4	2.6	2.5	2.6	2.6	2.8^	2.5	2.3*	19.3	22.7	23.0	22.6	24.4	24.3	24.7	23.8	22.9
0.2	0.2	1.9*	2.5	2.7^	2.6	2.7^	2.7^	2.7^	2.6	2.4	19.9	22.8	22.8	22.5	24.2	24.2	24.5	23.6	22.8
0.4	0.2	2.0*	2.4	2.6	2.5	2.6	2.6	2.7^	2.5	2.3*	20.0	22.4	22.2	21.8	23.3	23.4	23.5	22.8	21.9
0.6	0.2	2.2*	2.5	2.6	2.5	2.6	2.6	2.6	2.5	2.3*	20.2	21.9	21.4	21.1	22.5	22.7	22.8	22.0	21.2
0.8	0.2	2.3*	2.5	2.6	2.5	2.6	2.6	[99]	2.5	2.3*	20.7	21.6	20.6	20.3	21.8	22.1	[99]	21.3	20.4
0.4	0.4	1.3*	2.6	2.7^	2.7^	2.7^	2.7^	2.8^	2.6	2.4	17.9	25.7	25.5	25.3	26.8	26.9	27.2	26.3	25.1
0.6	0.4	1.4*	2.6	2.6	2.5	2.6	2.6	2.7^	2.5	2.3*	18.2	25.3	24.7	24.4	25.7	26.1	[99]	25.2	24.1
0	0.5	0.7*	2.5	2.6	2.6	2.7^	2.7^	2.9^	2.6	2.4	15.8	28.7	29.6	29.4	30.9	30.9	31.9	30.4	29.1
0.8	0.5	1.2*	2.7^	2.6	2.6	2.6	2.7^	[99]	2.6	2.3*	17.8	27.4	26.2	26.1	27.3	27.9	[99]	26.8	25.5
0.6	0.6	0.6*	2.5	2.5	2.5	2.6	2.5	2.7^	2.5	2.3*	15.7	31.0	30.3	30.4	31.7	32.5	32.8	31.1	29.6
0.8	0.8	0.2*	2.7^	2.6	2.7^	2.8^	2.8^	[99]	2.8^	2.5	11.7	41.6	39.9	41.0	40.9	43.1	[99]	40.4	38.0

Table 1.49. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤2, n_k~beta(2,3), p_C=0.1, δ_0=-0.05

Correlation Structure		**Type I error rate (%)**									**Power (%)**								
r_C	r_{CE}	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**
0	0	2.5	2.7^	2.9^	2.9^	2.8^	2.8^	2.8^	2.6	2.5	77.5	77.4	75.6	76.2	79.4	79.4	79.4	78.1	77.3
0.2	0	2.6	2.7^	2.9^	2.9^	2.8^	2.8^	2.8^	2.6	2.4	77.7	77.7	75.3	76.0	79.1	79.1	79.0	77.8	77.1
0.4	0	2.5	2.6	2.8^	2.8^	2.6	2.7^	2.6	2.4	2.3*	77.5	77.4	74.6	75.2	78.5	78.5	78.3	77.0	76.3
0.6	0	2.6	2.5	2.6	2.7^	2.6	2.6	[99]	2.4	2.3*	77.2	77.0	73.6	74.3	77.6	77.6	77.5	76.1	75.3
0.8	0	2.9^	2.6	2.5	2.6	2.6	2.5	[99]	2.4	2.3*	76.9	76.2	72.5	73.2	76.1	76.2	[99]	74.7	73.9
0	0.2	1.9*	2.6	2.7^	2.8^	2.7^	2.6	2.7^	2.5	2.3*	79.2	81.8	79.9	80.7	83.5	83.6	83.7	82.3	81.6
0.2	0.2	2.1*	2.7^	2.7^	2.8^	2.8^	2.8^	2.9^	2.6	2.4	78.9	81.3	78.9	79.7	82.7	82.7	82.7	81.5	80.8
0.4	0.2	2.0*	2.6	2.6	2.7^	2.7^	2.7^	2.7^	2.5	2.4	78.6	81.0	78.3	79.1	82.1	82.1	82.1	80.9	80.1
0.6	0.2	2.2*	2.6	2.8^	2.9^	2.7^	2.7^	2.6	2.5	2.3*	78.4	80.6	77.5	78.3	81.1	81.3	81.2	79.8	78.9
0.8	0.2	2.3*	2.5	2.5	2.6	2.5	2.5	[99]	2.4	2.2*	77.8	79.7	75.8	76.6	79.5	79.8	[99]	78.0	77.2
0.4	0.4	1.5*	2.7^	2.7^	2.9^	2.7^	2.8^	2.9^	2.5	2.4	80.5	85.8	83.7	84.5	86.8	86.9	87.0	85.7	85.0
0.6	0.4	1.6*	2.6	2.6	2.7^	2.6	2.6	[99]	2.4	2.3*	79.8	85.0	82.4	83.3	85.8	85.9	85.8	84.6	83.8
0	0.5	1.0*	2.7^	2.6	2.8^	2.7^	2.8^	2.9^	2.5	2.4	81.7	89.1	88.2	89.0	90.7	90.8	91.0	89.9	89.3
0.8	0.5	1.4*	2.7^	2.6	2.8^	2.7^	2.7^	[99]	2.5	2.3*	80.2	87.4	84.2	85.2	87.2	87.7	[99]	86.0	85.1
0.6	0.6	0.9*	2.7^	2.6	2.9^	2.8^	2.8^	2.9^	2.6	2.4	82.0	91.0	89.0	90.0	91.4	91.8	91.7	90.5	89.9
0.8	0.8	0.3*	3.0^	2.6	3.1^	3.0^	3.0^	[99]	2.8^	2.5	85.4	97.3	96.0	96.7	97.1	97.5	[99]	96.5	96.1

Table 1.50. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤2, n_k~beta(2,3), p_C=0.2, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.5	2.5	2.7^	2.6	2.6	2.6	2.7^	2.6	2.5	54.9	53.8	53.0	52.9	56.0	56.0	56.1	55.4	54.6
0.2	0	2.6	2.6	2.7^	2.7^	2.6	2.6	2.6	2.6	2.5	55.0	53.4	52.2	52.1	55.2	55.2	55.2	54.6	53.9
0.4	0	2.7^	2.5	2.6	2.6	2.6	2.6	2.6	2.5	2.4	54.2	52.7	50.9	50.8	53.5	53.5	53.5	52.8	52.1
0.6	0	2.9^	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.4	54.5	52.4	49.9	49.8	52.7	52.8	53.0	52.0	51.2
0.8	0	3.0^	2.5	2.4	2.4	2.4	2.4	2.5	2.4	2.3*	54.2	51.6	48.4	48.3	51.2	51.4	51.7	50.6	49.8
0	0.2	1.9*	2.6	2.7^	2.6	2.7^	2.7^	2.7^	2.6	2.5	55.5	59.0	58.4	58.4	61.5	61.5	61.6	60.8	60.1
0.2	0.2	1.9*	2.5	2.6	2.6	2.6	2.6	2.6	2.5	2.4	55.3	58.4	57.5	57.5	60.4	60.4	60.4	59.8	59.0
0.4	0.2	2.0*	2.5	2.6	2.6	2.5	2.5	2.5	2.5	2.3*	55.0	58.0	56.1	56.2	59.1	59.2	59.2	58.5	57.7
0.6	0.2	2.2*	2.5	2.6	2.6	2.5	2.6	2.6	2.5	2.4	55.2	57.8	55.4	55.5	58.1	58.4	58.5	57.4	56.5
0.8	0.2	2.2*	2.4	2.4	2.4	2.4	2.4	[99]	2.3*	2.2*	54.8	56.7	53.5	53.6	56.1	56.7	56.9	55.5	54.7
0.4	0.4	1.4*	2.5	2.7^	2.7^	2.7^	2.6	2.7^	2.6	2.5	55.4	64.5	63.0	63.2	65.7	66.1	66.1	65.1	64.2
0.6	0.4	1.5*	2.7^	2.7^	2.7^	2.7^	2.7^	2.8^	2.6	2.5	55.7	64.1	62.0	62.2	64.6	65.0	65.1	64.0	63.0
0	0.5	0.8*	2.5	2.7^	2.7^	2.6	2.6	2.8^	2.5	2.4	55.6	70.0	70.5	70.7	73.0	73.0	73.8	72.5	71.6
0.8	0.5	1.2*	2.4	2.6	2.6	2.5	2.5	2.5	2.4	2.3*	55.5	67.6	64.2	64.5	66.7	67.8	68.0	66.1	65.1
0.6	0.6	0.7*	2.5	2.6	2.6	2.7^	2.6	2.6	2.6	2.4	55.9	73.3	71.9	72.4	73.9	74.8	74.8	73.3	72.3
0.8	0.8	0.3*	2.6	2.6	2.8^	2.8^	2.7^	2.9^	2.8^	2.5	57.1	86.9	85.2	86.0	86.1	87.8	87.6	85.5	84.5

Table 1.51. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, $n_k \le 2$, n_k~beta(2,3), p_C=0.3, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.6	2.6	2.7^	2.6	2.7^	2.7^	2.7^	2.6	2.6	44.6	43.3	42.5	42.2	45.2	45.2	45.3	44.9	44.2
0.2	0	2.6	2.5	2.6	2.5	2.6	2.6	2.6	2.5	2.4	44.8	43.2	42.1	41.9	44.4	44.4	44.5	44.1	43.5
0.4	0	2.8^	2.5	2.6	2.6	2.6	2.6	2.6	2.6	2.5	44.8	42.5	41.2	40.9	43.4	43.4	43.4	43.1	42.5
0.6	0	2.9^	2.4	2.5	2.4	2.5	2.5	2.4	2.5	2.4	44.8	42.0	40.0	39.8	42.3	42.4	42.5	42.0	41.4
0.8	0	3.2^	2.5	2.5	2.5	2.6	2.6	2.6	2.6	2.4	45.1	41.8	39.2	38.9	41.4	41.6	42.1	41.1	40.4
0	0.2	1.7*	2.5	2.6	2.5	2.6	2.6	2.6	2.5	2.4	45.0	49.0	48.9	48.6	51.5	51.4	51.6	51.2	50.4
0.2	0.2	1.9*	2.5	2.7^	2.6	2.6	2.5	2.6	2.5	2.4	44.8	48.3	47.5	47.3	49.9	49.9	50.0	49.6	48.9
0.4	0.2	2.1*	2.6	2.7^	2.6	2.7^	2.6	2.7^	2.6	2.5	44.0	47.2	45.9	45.7	48.0	48.3	48.3	47.8	47.1
0.6	0.2	2.2*	2.5	2.5	2.5	2.6	2.6	2.6	2.5	2.4	44.6	47.0	44.9	44.7	47.3	47.5	47.7	47.0	46.3
0.8	0.2	2.2*	2.4	2.4	2.4	2.4	2.3*	2.4	2.3*	2.2*	44.8	46.5	44.1	43.9	45.9	46.5	46.9	45.6	44.9
0.4	0.4	1.2*	2.5	2.5	2.5	2.6	2.6	2.6	2.5	2.4	43.9	54.2	53.3	53.2	55.4	55.7	55.9	55.2	54.2
0.6	0.4	1.4*	2.6	2.5	2.4	2.6	2.5	2.6	2.6	2.4	43.9	53.2	51.7	51.6	53.5	54.2	54.3	53.3	52.4
0	0.5	0.7*	2.5	2.5	2.5	2.6	2.5	2.6	2.5	2.4	43.4	60.3	61.4	61.4	63.4	63.3	64.4	63.2	62.2
0.8	0.5	1.2*	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	44.4	57.5	54.6	54.6	56.5	57.9	58.1	56.3	55.3
0.6	0.6	0.7*	2.5	2.5	2.5	2.6	2.6	2.7^	2.6	2.4	43.5	63.2	61.7	61.9	63.5	64.4	64.5	63.2	62.2
0.8	0.8	0.2*	2.6	2.6	2.7^	2.7^	2.7^	2.8^	2.7^	2.5	42.3	78.9	77.1	77.7	77.2	79.9	80.0	77.0	75.8

Table 1.52. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k≤2, n_k~beta(2,3), p_C=0.1, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.7^	2.7^	2.7^	2.7^	2.7^	2.7^	2.5	2.5	97.3	97.1	96.5	96.6	97.6	97.6	97.6	97.4	97.3
0.2	0	2.6	2.7^	2.7^	2.8^	2.7^	2.7^	2.7^	2.6	2.5	97.2	97.1	96.2	96.3	97.4	97.4	97.4	97.2	97.1
0.4	0	2.7^	2.6	2.7^	2.8^	2.7^	2.7^	2.6	2.6	2.5	97.2	97.0	96.1	96.2	97.3	97.3	97.2	97.1	97.0
0.6	0	2.7^	2.5	2.5	2.6	2.5	2.6	2.6	2.4	2.3*	96.9	96.7	95.6	95.8	96.8	96.8	96.8	96.6	96.4
0.8	0	2.9^	2.6	2.6	2.7^	2.5	2.5	2.6	2.4	2.3*	96.8	96.5	95.1	95.2	96.6	96.6	96.7	96.3	96.1
0	0.2	1.9*	2.6	2.6	2.7^	2.7^	2.7^	2.7^	2.6	2.4	97.9	98.1	97.6	97.7	98.5	98.4	98.4	98.3	98.3
0.2	0.2	1.9*	2.6	2.6	2.7^	2.6	2.6	2.6	2.4	2.4	97.7	98.0	97.4	97.5	98.3	98.3	98.3	98.2	98.1
0.4	0.2	2.2*	2.7^	2.7^	2.8^	2.8^	2.8^	2.7^	2.6	2.5	97.6	98.0	97.2	97.3	98.2	98.2	98.2	98.0	97.9
0.6	0.2	2.2*	2.6	2.7^	2.7^	2.6	2.6	2.6	2.5	2.4	97.5	97.8	96.8	97.0	97.9	97.9	97.9	97.7	97.6
0.8	0.2	2.3*	2.5	2.4	2.5	2.5	2.5	2.5	2.4	2.3*	97.2	97.6	96.3	96.5	97.5	97.6	97.6	97.3	97.1
0.4	0.4	1.5*	2.6	2.8^	2.9^	2.8^	2.8^	2.8^	2.6	2.5	98.3	99.0	98.7	98.8	99.1	99.2	99.2	99.0	99.0
0.6	0.4	1.5*	2.5	2.5	2.6	2.6	2.5	2.6	2.4	2.3*	98.1	98.9	98.3	98.4	98.9	99.0	98.9	98.8	98.7
0	0.5	1.0*	2.6	2.7^	2.8^	2.7^	2.7^	2.8^	2.5	2.4	98.8	99.5	99.3	99.4	99.6	99.6	99.6	99.6	99.5
0.8	0.5	1.4*	2.6	2.5	2.7^	2.6	2.7^	2.7^	2.5	2.4	98.3	99.2	98.7	98.8	99.2	99.2	99.2	99.1	99.0
0.6	0.6	0.9*	2.6	2.6	2.8^	2.6	2.6	2.7^	2.5	2.4	98.9	99.7	99.5	99.6	99.7	99.7	99.7	99.7	99.6
0.8	0.8	0.3*	2.9^	2.5	2.9^	2.9^	2.9^	3.0^	2.7^	2.5	99.6	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0

Table 1.53. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k≤2, n_k~beta(2,3), p_C=0.2, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	84.1	82.8	81.5	81.5	84.6	84.5	84.6	84.3	84.0
0.2	0	2.7^	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.6	2.6	83.5	82.2	80.8	80.8	83.4	83.4	83.4	83.1	82.8
0.4	0	2.6	2.5	2.6	2.6	2.5	2.5	2.5	2.4	2.3*	83.4	82.0	79.7	79.7	82.7	82.7	82.7	82.4	82.1
0.6	0	2.9^	2.5	2.6	2.6	2.5	2.5	2.5	2.5	2.4	83.0	81.7	79.1	79.1	81.8	81.9	82.0	81.5	81.1
0.8	0	3.1^	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.5	82.6	80.8	77.7	77.7	80.4	80.6	81.0	80.1	79.7
0	0.2	1.8*	2.6	2.6	2.6	2.6	2.6	2.7^	2.5	2.5	85.3	86.8	86.0	86.1	88.3	88.3	88.4	88.2	87.9
0.2	0.2	1.8*	2.5	2.6	2.6	2.5	2.5	2.6	2.5	2.4	85.2	86.6	85.6	85.6	87.9	87.9	87.9	87.7	87.3
0.4	0.2	2.1*	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.4	85.0	86.5	85.1	85.2	87.2	87.3	87.4	87.0	86.7
0.6	0.2	2.2*	2.6	2.5	2.5	2.5	2.6	2.6	2.5	2.4	84.4	85.7	83.7	83.7	86.0	86.1	86.2	85.7	85.4
0.8	0.2	2.3*	2.5	2.5	2.5	2.4	2.4	2.5	2.4	2.3*	84.1	85.2	82.4	82.4	84.7	85.2	85.4	84.4	84.0
0.4	0.4	1.3*	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.4	86.5	90.8	90.0	90.0	91.4	91.7	91.7	91.3	90.9
0.6	0.4	1.5*	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	86.3	90.6	89.0	89.1	90.7	91.1	91.1	90.5	90.2
0	0.5	0.8*	2.5	2.6	2.6	2.5	2.5	2.7^	2.5	2.4	88.9	94.1	94.4	94.5	95.3	95.4	95.6	95.2	95.1
0.8	0.5	1.3*	2.6	2.5	2.6	2.6	2.6	2.6	2.5	2.4	87.1	92.6	90.8	91.0	92.1	92.7	92.8	91.9	91.6
0.6	0.6	0.7*	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.4	88.9	95.5	94.8	94.9	95.6	95.9	95.9	95.5	95.3
0.8	0.8	0.2*	2.6	2.5	2.7^	2.5	2.6	2.7^	2.5	2.3*	92.4	99.2	98.8	98.9	98.9	99.2	99.3	98.9	98.8

Table 1.54. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, $n_k \le 2$, $n_k \sim beta(2,3)$, $p_C=0.3$, $\delta_0=-0.05$

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.6	2.6	2.6	2.5	2.5	2.5	2.5	2.5	73.5	71.8	70.7	70.6	73.8	73.8	73.8	73.7	73.3
0.2	0	2.7^	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.5	73.3	71.5	69.9	69.7	72.8	72.9	72.9	72.7	72.3
0.4	0	2.8^	2.6	2.6	2.5	2.5	2.5	2.5	2.5	2.5	73.3	71.1	69.1	68.9	72.0	72.1	72.1	71.9	71.5
0.6	0	3.0^	2.5	2.6	2.6	2.5	2.5	2.5	2.5	2.5	73.0	70.6	68.1	67.9	70.9	71.1	71.2	70.8	70.3
0.8	0	3.1^	2.6	2.5	2.5	2.5	2.5	2.5	2.5	2.4	72.2	69.5	66.2	66.0	68.9	69.2	69.7	68.8	68.4
0	0.2	1.7*	2.5	2.5	2.5	2.5	2.5	2.5	2.4	2.4	74.5	77.2	76.7	76.7	79.2	79.2	79.3	79.1	78.7
0.2	0.2	1.9*	2.5	2.6	2.6	2.6	2.6	2.6	2.6	2.5	74.8	77.0	76.1	76.0	78.6	78.7	78.7	78.5	78.1
0.4	0.2	2.1*	2.6	2.7^	2.6	2.6	2.6	2.6	2.6	2.5	74.6	76.6	75.0	74.9	77.7	77.8	77.8	77.6	77.1
0.6	0.2	2.1*	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.4	74.1	76.1	73.8	73.7	76.3	76.5	76.7	76.2	75.7
0.8	0.2	2.5	2.6	2.7^	2.7^	2.6	2.7^	2.7^	2.6	2.6	73.5	75.2	71.9	71.8	74.4	75.1	75.6	74.3	73.8
0.4	0.4	1.2*	2.5	2.6	2.6	2.6	2.6	2.6	2.6	2.5	76.4	83.3	82.4	82.4	84.3	84.5	84.6	84.2	83.8
0.6	0.4	1.3*	2.5	2.5	2.5	2.4	2.5	2.5	2.4	2.4	76.1	82.8	81.2	81.2	83.0	83.6	83.5	82.9	82.5
0	0.5	0.7*	2.5	2.6	2.6	2.6	2.6	2.6	2.5	2.4	78.5	88.1	88.9	88.9	90.1	90.2	90.6	90.1	89.8
0.8	0.5	1.2*	2.6	2.5	2.5	2.5	2.5	2.6	2.5	2.4	76.6	85.5	82.9	83.0	84.6	85.5	85.8	84.5	84.0
0.6	0.6	0.7*	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.4	78.1	90.2	89.2	89.3	90.2	90.9	91.0	90.1	89.8
0.8	0.8	0.2*	2.5	2.5	2.6	2.5	2.6	2.6	2.6	2.4	82.3	97.3	96.9	97.1	96.9	97.6	97.7	96.8	96.6

Table 1.55. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k≤2, n_k~beta(2,3), p_C=0.05, δ_0=-0.02

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.5	2.9^	2.8^	3.0^	2.8^	2.8^	2.9^	2.6	2.5	56.5	57.2	55.4	56.4	59.2	59.2	59.3	57.3	56.5
0.2	0	2.5	2.6	2.8^	2.9^	2.8^	2.8^	2.8^	2.5	2.4	56.2	57.0	55.0	56.1	58.7	58.6	58.6	56.7	55.8
0.4	0	2.6	2.7^	2.7^	2.9^	2.8^	2.8^	2.7^	2.6	2.4	55.9	56.3	53.9	54.9	57.6	57.6	57.4	55.7	54.7
0.6	0	2.6	2.6	2.7^	2.8^	2.6	2.6	2.6	2.4	2.3*	56.1	56.2	53.2	54.3	56.9	56.9	56.7	54.9	54.0
0.8	0	2.8^	2.6	2.6	2.7^	2.6	2.6	[99]	2.4	2.2*	55.9	55.7	51.7	52.8	55.3	55.5	[99]	53.4	52.3
0	0.2	2.0*	2.6	2.6	2.8^	2.7^	2.7^	2.7^	2.4	2.3*	56.3	59.6	57.6	58.7	61.7	61.7	61.9	59.8	58.8
0.2	0.2	2.1*	2.7^	2.7^	2.9^	2.7^	2.7^	2.7^	2.5	2.4	56.7	59.9	57.9	59.0	61.7	61.6	61.7	59.7	58.7
0.4	0.2	2.2*	2.8^	2.8^	2.9^	2.8^	2.8^	2.8^	2.6	2.4	56.3	59.2	56.5	57.6	60.5	60.5	60.5	58.6	57.7
0.6	0.2	2.3*	2.7^	2.7^	2.8^	2.8^	2.8^	2.7^	2.5	2.4	56.1	58.8	55.5	56.8	59.4	59.5	59.5	57.4	56.3
0.8	0.2	2.6	2.8^	2.7^	2.8^	2.8^	2.8^	[99]	2.6	2.5	56.1	58.0	53.8	55.2	57.9	58.1	[99]	55.9	54.8
0.4	0.4	1.6*	2.6	2.7^	2.9^	2.7^	2.7^	2.7^	2.5	2.4	56.8	64.0	61.3	62.7	65.6	65.6	65.6	63.6	62.5
0.6	0.4	1.7*	2.7^	2.5	2.7^	2.7^	2.7^	2.7^	2.4	2.3*	56.5	63.0	60.1	61.4	63.6	63.8	63.8	61.7	60.6
0	0.5	1.1*	2.6	2.7^	2.8^	2.7^	2.7^	2.7^	2.4	2.2*	56.9	67.6	66.2	67.5	70.0	70.0	70.4	68.0	67.0
0.8	0.5	1.6*	2.8^	2.7^	2.9^	2.8^	2.7^	[99]	2.6	2.4	56.7	65.8	61.6	63.2	65.6	66.1	[99]	63.4	62.3
0.6	0.6	1.0*	2.7^	2.6	2.9^	2.8^	2.8^	2.9^	2.5	2.3*	56.9	70.6	67.6	69.2	71.3	71.8	71.8	69.3	68.0
0.8	0.8	0.4*	2.8^	2.6	3.0^	2.8^	2.8^	[99]	2.5	2.3*	58.7	83.4	80.2	82.2	82.8	84.1	[99]	81.0	79.8

Table 1.56. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤2, n_k~beta(2,3), p_C=0.05, δ_0=-0.02

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.4	2.8^	2.9^	3.1^	3.0^	3.0^	3.0^	2.6	2.4	31.1	33.2	32.1	33.2	34.9	34.9	35.0	32.3	31.1
0.2	0	2.4	2.8^	2.8^	3.0^	2.8^	2.8^	2.9^	2.5	2.3*	31.0	33.1	31.8	33.1	34.4	34.4	34.4	31.8	30.8
0.4	0	2.6	2.9^	3.0^	3.2^	2.9^	2.9^	[99]	2.6	2.4	30.9	32.6	30.7	32.0	33.6	33.5	33.4	31.0	29.7
0.6	0	2.7^	2.8^	2.9^	3.1^	2.9^	2.8^	[99]	2.5	2.4	31.8	32.8	30.7	32.1	33.6	33.5	[99]	31.0	29.8
0.8	0	2.8^	2.7^	2.6	2.8^	2.7^	2.7^	[99]	2.4	2.2*	31.8	32.0	29.4	30.7	32.2	32.1	[99]	29.7	28.5
0	0.2	2.0*	2.8^	3.0^	3.2^	3.0^	3.0^	3.0^	2.6	2.4	30.6	35.0	33.8	35.2	36.8	36.8	37.0	34.2	32.9
0.2	0.2	2.0*	2.7^	2.8^	3.0^	2.9^	2.9^	2.9^	2.5	2.3*	30.9	34.9	33.6	35.0	36.5	36.5	36.6	33.9	32.7
0.4	0.2	2.1*	2.8^	2.9^	3.1^	2.9^	2.9^	2.9^	2.5	2.3*	30.9	34.9	33.2	34.7	36.0	36.0	[99]	33.3	32.1
0.6	0.2	2.3*	2.8^	2.9^	3.1^	2.9^	2.9^	[99]	2.5	2.4	30.7	34.1	31.8	33.4	34.8	34.7	[99]	32.1	30.7
0.8	0.2	2.5	2.8^	2.7^	2.9^	2.8^	2.8^	[99]	2.5	2.2*	31.4	34.1	30.8	32.4	33.9	34.1	[99]	31.3	30.0
0.4	0.4	1.6*	2.9^	2.9^	3.1^	3.0^	3.0^	3.0^	2.6	2.4	30.3	38.3	36.5	38.3	39.6	39.6	39.6	36.7	35.3
0.6	0.4	1.7*	2.7^	2.8^	3.0^	2.8^	2.8^	[99]	2.5	2.3*	30.3	37.8	35.3	37.1	38.3	38.4	[99]	35.6	34.2
0	0.5	1.2*	2.9^	2.8^	3.1^	3.0^	3.0^	3.1^	2.6	2.3*	29.0	40.9	39.5	41.6	43.2	43.2	43.7	40.2	38.7
0.8	0.5	1.5*	2.8^	2.6	2.9^	2.8^	2.8^	[99]	2.4	2.2*	29.7	38.9	35.6	37.9	39.0	39.3	[99]	36.0	34.3
0.6	0.6	1.1*	2.9^	2.8^	3.2^	3.0^	2.9^	[99]	2.6	2.3*	28.6	43.3	40.4	42.9	44.2	44.4	[99]	41.0	39.2
0.8	0.8	0.4*	3.1^	2.7^	3.3^	3.1^	3.1^	[99]	2.7^	2.4	25.9	55.2	50.3	54.6	55.0	56.1	[99]	51.2	48.5

Table 1.57. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, $n_k \leq 2$, n_k~beta(2,3), p_C=0.2, δ_0=-0.02

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.5	2.7^	2.7^	2.7^	2.7^	2.7^	2.6	2.4	12.9	12.9	13.0	13.0	13.5	13.5	13.6	13.2	12.9
0.2	0	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.5	12.9	12.5	12.7	12.6	13.1	13.1	13.1	12.7	12.4
0.4	0	2.7^	2.5	2.6	2.5	2.6	2.6	2.6	2.5	2.4	13.5	12.7	12.6	12.6	13.2	13.1	13.1	12.9	12.4
0.6	0	2.8^	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.4	13.7	12.5	12.1	12.0	12.7	12.7	12.7	12.4	12.0
0.8	0	3.2^	2.5	2.6	2.6	2.6	2.6	2.6	2.5	2.4	14.1	12.3	11.7	11.7	12.3	12.4	[99]	12.0	11.7
0	0.2	1.8*	2.6	2.7^	2.7^	2.7^	2.7^	2.7^	2.6	2.5	11.7	14.1	14.4	14.4	15.0	15.0	15.1	14.7	14.2
0.2	0.2	2.0*	2.7^	2.7^	2.7^	2.7^	2.7^	2.7^	2.6	2.5	11.6	13.8	13.9	13.9	14.4	14.4	14.4	14.0	13.6
0.4	0.2	2.0*	2.5	2.6	2.6	2.6	2.6	2.6	2.5	2.4	11.7	13.5	13.4	13.3	13.8	13.9	13.9	13.5	13.1
0.6	0.2	2.3*	2.6	2.6	2.6	2.7^	2.6	2.6	2.6	2.4	12.4	13.6	13.4	13.4	13.9	14.0	14.0	13.6	13.2
0.8	0.2	2.4	2.6	2.6	2.6	2.6	2.6	[99]	2.5	2.4	12.8	13.3	13.1	13.1	13.4	13.5	13.4	13.1	12.6
0.4	0.4	1.3*	2.5	2.6	2.6	2.6	2.6	2.6	2.5	2.3*	10.1	15.4	15.2	15.2	15.7	15.8	15.9	15.4	14.8
0.6	0.4	1.5*	2.4	2.6	2.6	2.6	2.6	2.6	2.5	2.4	10.6	15.1	14.8	14.8	15.4	15.5	15.5	15.0	14.5
0	0.5	0.8*	2.6	2.7^	2.7^	2.7^	2.7^	2.8^	2.6	2.5	8.3	17.0	17.4	17.4	18.2	18.1	18.7	17.7	17.1
0.8	0.5	1.3*	2.6	2.6	2.6	2.5	2.6	2.6	2.5	2.3*	10.0	16.2	15.5	15.6	16.0	16.4	16.3	15.7	15.2
0.6	0.6	0.7*	2.5	2.6	2.6	2.6	2.6	2.6	2.5	2.4	8.1	17.6	17.4	17.5	17.9	18.3	18.4	17.5	16.7
0.8	0.8	0.2*	2.6	2.5	2.6	2.6	2.6	2.7^	2.6	2.3*	5.1	23.6	22.9	23.3	23.0	24.1	24.2	22.6	21.4

Table 1.58. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, $n_k \leq 2$, n_k~beta(2,3), p_C=0.2, δ_0=-0.1

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.5	2.6	2.6	2.5	2.5	2.5	2.5	2.4	98.6	98.4	98.0	98.0	98.7	98.7	98.7	98.6	98.6
0.2	0	2.6	2.5	2.7^	2.6	2.6	2.6	2.6	2.6	2.5	98.5	98.3	97.7	97.7	98.5	98.5	98.6	98.5	98.4
0.4	0	2.7^	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.3*	98.5	98.2	97.6	97.6	98.4	98.4	98.5	98.3	98.2
0.6	0	2.9^	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.4	98.3	98.0	97.3	97.3	98.1	98.2	98.2	98.1	98.0
0.8	0	3.1^	2.5	2.7^	2.6	2.5	2.5	2.6	2.5	2.4	98.2	97.9	96.9	96.9	97.8	97.9	97.9	97.7	97.6
0	0.2	1.8*	2.5	2.6	2.5	2.5	2.5	2.6	2.5	2.4	99.1	99.2	99.0	99.0	99.4	99.4	99.4	99.4	99.3
0.2	0.2	1.9*	2.5	2.7^	2.6	2.6	2.6	2.6	2.5	2.5	99.0	99.1	98.8	98.8	99.3	99.3	99.3	99.3	99.2
0.4	0.2	2.0*	2.5	2.6	2.6	2.6	2.6	2.6	2.6	2.4	98.9	99.1	98.7	98.8	99.2	99.2	99.3	99.1	99.1
0.6	0.2	2.2*	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	98.8	99.0	98.6	98.6	99.0	99.1	99.1	99.0	98.9
0.8	0.2	2.3*	2.5	2.6	2.6	2.5	2.6	2.6	2.5	2.4	98.7	98.9	98.3	98.3	98.8	98.9	98.9	98.7	98.7
0.4	0.4	1.3*	2.5	2.6	2.6	2.6	2.6	2.6	2.5	2.4	99.3	99.6	99.5	99.5	99.7	99.7	99.7	99.7	99.6
0.6	0.4	1.4*	2.6	2.6	2.6	2.6	2.6	2.7^	2.6	2.4	99.2	99.6	99.4	99.4	99.6	99.6	99.6	99.6	99.5
0	0.5	0.8*	2.5	2.5	2.6	2.6	2.6	2.7^	2.5	2.4	99.7	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
0.8	0.5	1.2*	2.7^	2.5	2.6	2.6	2.7^	2.7^	2.6	2.5	99.4	99.8	99.6	99.6	99.7	99.8	[99]	99.7	99.7
0.6	0.6	0.7*	2.6	2.5	2.7^	2.7^	2.7^	2.8^	2.6	2.5	99.7	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
0.8	0.8	0.2*	2.8^	2.5	2.9^	2.8^	2.8^	3.1^	2.8^	2.6	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.59. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤2, n_k~beta(2,3), p_C=0.3, δ_0=-0.1

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.4	2.5	2.4	2.4	2.4	2.5	2.4	2.3*	95.7	94.9	94.4	94.3	95.7	95.7	95.7	95.7	95.6
0.2	0	2.6	2.4	2.6	2.5	2.5	2.5	2.5	2.5	2.4	95.3	94.6	93.7	93.6	95.1	95.1	95.2	95.1	94.9
0.4	0	2.9^	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	95.1	94.3	93.4	93.3	94.7	94.7	94.8	94.7	94.5
0.6	0	2.8^	2.3*	2.4	2.3*	2.4	2.4	2.4	2.4	2.3*	95.0	94.1	92.9	92.7	94.3	94.4	94.4	94.3	94.1
0.8	0	3.2^	2.4	2.6	2.5	2.4	2.5	2.5	2.4	2.3*	94.6	93.7	91.8	91.7	93.4	93.6	93.9	93.4	93.2
0	0.2	1.7*	2.4	2.5	2.4	2.5	2.5	2.5	2.5	2.5	96.5	96.9	96.7	96.6	97.5	97.5	97.5	97.5	97.4
0.2	0.2	1.9*	2.4	2.5	2.4	2.5	2.5	2.5	2.5	2.4	96.4	96.8	96.6	96.6	97.3	97.3	97.3	97.3	97.2
0.4	0.2	2.0*	2.4	2.6	2.5	2.6	2.5	2.6	2.6	2.5	96.2	96.7	96.0	95.9	97.0	97.0	97.1	97.0	96.8
0.6	0.2	2.1*	2.4	2.5	2.4	2.4	2.5	2.5	2.5	2.4	96.1	96.5	95.6	95.5	96.6	96.7	96.7	96.6	96.4
0.8	0.2	2.3*	2.5	2.6	2.5	2.5	2.5	2.5	2.5	2.4	95.8	96.3	94.9	94.9	96.1	96.3	96.4	96.1	95.9
0.4	0.4	1.2*	2.4	2.5	2.5	2.5	2.4	2.5	2.5	2.4	97.3	98.4	98.2	98.2	98.6	98.7	98.7	98.6	98.5
0.6	0.4	1.4*	2.5	2.6	2.6	2.5	2.6	2.6	2.5	2.4	97.0	98.4	97.8	97.9	98.4	98.5	98.6	98.4	98.3
0	0.5	0.7*	2.5	2.5	2.5	2.6	2.6	2.7^	2.6	2.4	98.4	99.3	99.4	99.4	99.6	99.6	99.6	99.6	99.6
0.8	0.5	1.1*	2.5	2.5	2.5	2.6	2.6	2.6	2.6	2.5	97.6	99.0	98.5	98.5	98.9	99.0	99.1	98.9	98.8
0.6	0.6	0.6*	2.4	2.5	2.6	2.5	2.5	2.6	2.5	2.4	98.4	99.6	99.5	99.5	99.6	99.6	99.6	99.6	99.5
0.8	0.8	0.2*	2.6	2.4	2.7^	2.7^	2.7^	2.8^	2.8^	2.6	99.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.60. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤2, n_k~beta(2,3), p_C=0.5, δ_0=-0.1

Correlation Structure		**Type I error rate (%)**									**Power (%)**								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.6	2.5	2.6	2.5	2.5	2.5	2.6	2.6	2.5	92.5	91.0	90.7	90.4	92.4	92.4	92.4	92.5	92.3
0.2	0	2.7^	2.4	2.5	2.4	2.4	2.4	2.5	2.4	2.3*	91.9	90.6	89.8	89.6	91.5	91.5	91.6	91.6	91.3
0.4	0	2.7^	2.4	2.5	2.4	2.4	2.5	2.5	2.5	2.4	91.7	90.2	88.9	88.7	90.7	90.8	90.9	90.9	90.6
0.6	0	2.9^	2.3*	2.5	2.4	2.4	2.4	2.4	2.4	2.4	91.0	89.5	87.9	87.6	89.6	89.7	90.0	89.7	89.4
0.8	0	3.3^	2.5	2.5	2.4	2.6	2.6	2.6	2.6	2.5	91.2	89.6	87.3	87.0	89.2	89.4	89.9	89.3	89.0
0	0.2	1.8*	2.5	2.5	2.5	2.6	2.6	2.6	2.6	2.5	93.8	94.3	94.3	94.1	95.4	95.4	95.5	95.5	95.4
0.2	0.2	1.8*	2.4	2.5	2.4	2.4	2.4	2.5	2.5	2.3*	93.4	94.0	93.7	93.6	94.8	94.8	94.8	94.9	94.7
0.4	0.2	2.0*	2.4	2.5	2.4	2.5	2.5	2.5	2.5	2.4	93.3	93.8	93.1	92.9	94.3	94.4	94.5	94.4	94.2
0.6	0.2	2.2*	2.5	2.5	2.4	2.5	2.5	2.6	2.5	2.4	92.9	93.5	92.4	92.2	93.6	93.9	93.9	93.7	93.5
0.8	0.2	2.4	2.4	2.5	2.4	2.5	2.5	2.5	2.5	2.4	92.5	93.2	91.4	91.3	92.7	93.2	93.4	92.8	92.5
0.4	0.4	1.2*	2.4	2.5	2.5	2.5	2.5	2.6	2.5	2.4	94.9	96.9	96.7	96.6	97.2	97.4	97.4	97.3	97.2
0.6	0.4	1.4*	2.5	2.5	2.4	2.5	2.5	2.5	2.6	2.5	94.6	96.8	96.2	96.2	96.7	97.0	97.1	96.8	96.6
0	0.5	0.7*	2.4	2.5	2.4	2.5	2.4	2.6	2.5	2.4	96.4	98.4	98.7	98.7	99.0	99.0	99.1	99.0	98.9
0.8	0.5	1.2*	2.4	2.6	2.6	2.6	2.6	2.6	2.6	2.5	95.2	97.9	97.1	97.1	97.6	97.9	98.0	97.6	97.5
0.6	0.6	0.7*	2.4	2.5	2.5	2.6	2.6	2.6	2.6	2.5	96.4	98.9	98.7	98.8	98.9	99.1	99.1	99.0	98.9
0.8	0.8	0.2*	2.6	2.5	2.7^	2.6	2.7^	2.9^	2.7^	2.5	98.3	99.9	99.9	99.9	99.9	100.0	100.0	99.9	99.9

Table 1.61. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k=5, n_k~constant, p_C=0.1, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.2*	3.7^	3.6^	3.7^	3.8^	3.8^	4.1^	2.4	2.3*	47.1	56.8	56.0	56.6	57.5	57.5	58.1	48.7	48.1
0.2	0	2.5	3.6^	3.5^	3.6^	3.7^	3.7^	3.8^	2.4	2.3*	47.4	55.1	54.3	54.9	55.8	55.8	55.5	47.0	46.2
0.4	0	2.6	3.3^	3.1^	3.2^	3.4^	3.4^	3.4^	2.2*	2.1*	47.6	52.9	52.0	52.7	53.7	53.7	52.3	45.1	44.1
0.6	0	3.1^	3.3^	3.2^	3.3^	3.4^	3.4^	3.2^	2.2*	2.0*	47.4	50.0	48.9	49.8	50.7	50.7	[99]	42.1	40.9
0.8	0	3.5^	3.0^	2.8^	2.9^	3.1^	3.1^	[99]	2.1*	1.9*	47.8	46.7	45.5	46.5	47.4	47.4	[99]	39.2	37.8
0	0.2	1.6*	3.7^	3.6^	3.7^	3.8^	3.8^	4.6^	2.3*	2.2*	47.4	62.6	61.6	62.3	63.2	63.2	65.8	53.8	53.1
0.2	0.2	1.7*	3.6^	3.5^	3.6^	3.7^	3.7^	4.2^	2.3*	2.1*	47.6	60.7	59.7	60.5	61.3	61.3	63.0	52.2	51.2
0.4	0.2	2.0*	3.5^	3.3^	3.4^	3.6^	3.6^	3.8^	2.2*	2.1*	47.2	57.6	56.4	57.3	58.2	58.2	58.7	49.0	47.9
0.6	0.2	2.3*	3.3^	3.2^	3.3^	3.5^	3.5^	[99]	2.1*	1.9*	47.5	54.4	53.1	54.2	55.1	55.1	54.4	46.1	44.7
0.8	0.2	2.8^	3.1^	2.9^	3.0^	3.2^	3.2^	[99]	2.0*	1.8*	47.4	50.2	48.7	50.0	50.9	50.9	[99]	42.1	40.6
0.4	0.4	1.2*	3.4^	3.2^	3.4^	3.6^	3.6^	4.5^	2.1*	1.9*	46.9	65.1	63.8	64.9	65.7	65.7	68.2	56.0	54.6
0.6	0.4	1.5*	3.2^	3.0^	3.2^	3.3^	3.3^	[99]	2.0*	1.8*	47.1	61.0	59.4	60.8	61.7	61.7	[99]	52.2	50.6
0	0.5	0.8*	3.6^	3.3^	3.5^	3.7^	3.7^	5.5^	2.1*	1.9*	46.9	73.6	72.4	73.5	74.2	74.2	79.9	64.8	63.2
0.8	0.5	1.4*	3.1^	2.8^	3.1^	3.3^	3.3^	[99]	2.0*	1.7*	47.3	60.6	58.6	60.4	61.2	61.2	[99]	51.8	49.7
0.6	0.6	0.7*	3.6^	3.2^	3.6^	3.8^	3.8^	5.3^	2.1*	1.8*	46.5	72.3	70.4	72.2	73.0	73.0	76.7	63.3	61.0
0.8	0.8	0.2*	3.7^	2.9^	3.7^	3.9^	3.9^	[99]	2.0*	1.6*	46.5	85.0	82.3	84.8	85.4	85.4	[99]	77.0	74.1

Table 1.62. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k=5, n_k~constant, p_C=0.2, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	3.1^	3.1^	3.1^	3.2^	3.2^	3.4^	2.5	2.5	33.1	37.5	37.7	37.3	38.1	38.1	38.9	33.6	33.2
0.2	0	2.7^	2.9^	2.9^	2.9^	3.0^	3.0^	3.2^	2.4	2.3*	33.3	35.2	35.4	35.0	35.9	35.9	35.8	31.4	30.8
0.4	0	2.9^	2.8^	2.8^	2.7^	2.9^	2.9^	2.9^	2.3*	2.2*	34.0	33.4	33.5	33.2	34.0	34.0	33.4	29.9	29.2
0.6	0	3.4^	2.7^	2.7^	2.6	2.8^	2.8^	2.8^	2.2*	2.0*	34.1	30.7	30.8	30.4	31.3	31.3	30.2	27.4	26.6
0.8	0	4.1^	2.6	2.7^	2.6	2.8^	2.8^	[99]	2.2*	2.0*	34.7	28.3	28.3	28.1	28.9	28.9	[99]	25.2	24.3
0	0.2	1.5*	3.1^	3.1^	3.0^	3.2^	3.2^	3.8^	2.4	2.3*	31.3	43.0	43.1	42.8	43.7	43.7	46.1	38.4	37.8
0.2	0.2	1.8*	2.9^	2.9^	2.9^	3.0^	3.0^	3.5^	2.3*	2.2*	32.0	40.6	40.7	40.5	41.3	41.3	42.9	36.3	35.6
0.4	0.2	2.2*	3.0^	3.0^	3.0^	3.1^	3.1^	3.4^	2.4	2.2*	32.5	37.9	37.9	37.7	38.5	38.5	39.1	33.8	32.9
0.6	0.2	2.6	2.8^	2.8^	2.8^	3.0^	3.0^	3.0^	2.2*	2.1*	33.6	35.3	35.3	35.1	36.0	36.0	36.0	31.6	30.6
0.8	0.2	3.1^	2.6	2.6	2.6	2.8^	2.8^	[99]	2.1*	2.0*	33.8	31.7	31.7	31.6	32.4	32.4	[99]	28.4	27.3
0.4	0.4	1.1*	3.0^	3.0^	2.9^	3.1^	3.1^	3.5^	2.3*	2.1*	30.7	45.3	45.1	45.1	45.9	45.9	48.2	40.5	39.3
0.6	0.4	1.5*	2.9^	2.9^	2.8^	3.0^	3.0^	3.4^	2.2*	2.0*	31.4	41.1	40.9	40.9	41.8	41.8	43.2	36.8	35.4
0	0.5	0.5*	3.0^	3.0^	3.0^	3.1^	3.1^	4.3^	2.2*	2.0*	28.1	56.0	55.8	55.7	56.7	56.7	62.6	50.5	49.0
0.8	0.5	1.5*	2.7^	2.7^	2.7^	2.9^	2.9^	3.3^	2.2*	2.0*	31.4	41.0	40.5	40.8	41.7	41.7	[99]	36.6	34.9
0.6	0.6	0.6*	3.0^	2.9^	3.0^	3.1^	3.1^	3.9^	2.2*	1.9*	28.6	52.7	52.1	52.5	53.4	53.4	56.7	47.4	45.4
0.8	0.8	0.1*	3.0^	2.8^	3.0^	3.2^	3.2^	4.7^	2.2*	1.7*	25.0	67.0	65.5	66.8	67.6	67.6	71.9	61.2	58.1

Table 1.63. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k=5, n_k~constant, p_C=0.3, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.5	2.9^	3.0^	2.9^	3.0^	3.0^	3.3^	2.6	2.5	27.1	29.4	29.8	29.2	30.0	30.0	30.7	27.5	27.1
0.2	0	2.9^	2.8^	2.8^	2.7^	2.9^	2.9^	3.0^	2.5	2.4	28.1	27.7	28.0	27.5	28.3	28.3	28.7	26.0	25.6
0.4	0	3.4^	2.7^	2.8^	2.7^	2.8^	2.8^	3.0^	2.5	2.4	28.8	26.0	26.3	25.9	26.6	26.6	26.6	24.4	23.8
0.6	0	4.0^	2.7^	2.8^	2.7^	2.9^	2.9^	2.9^	2.5	2.4	29.5	24.0	24.3	23.9	24.6	24.6	24.3	22.6	22.0
0.8	0	4.6^	2.6	2.7^	2.6	2.7^	2.7^	2.6	2.3*	2.2*	30.3	22.3	22.5	22.2	22.8	22.8	22.2	21.0	20.3
0	0.2	1.5*	2.9^	2.9^	2.8^	3.0^	3.0^	3.4^	2.5	2.4	25.1	35.1	35.5	35.0	35.8	35.8	37.8	32.9	32.3
0.2	0.2	1.8*	2.7^	2.8^	2.7^	2.8^	2.8^	3.1^	2.5	2.3*	25.9	32.5	32.8	32.3	33.1	33.1	34.2	30.4	29.7
0.4	0.2	2.3*	2.8^	2.8^	2.7^	2.9^	2.9^	3.1^	2.5	2.4	27.0	30.3	30.6	30.1	30.9	30.9	31.5	28.3	27.5
0.6	0.2	2.6	2.5	2.6	2.5	2.7^	2.7^	2.8^	2.3*	2.1*	27.8	27.8	28.0	27.6	28.4	28.4	28.8	26.1	25.2
0.8	0.2	3.3^	2.5	2.5	2.5	2.6	2.6	2.7^	2.3*	2.1*	28.8	25.3	25.5	25.1	25.8	25.8	25.9	23.8	22.9
0.4	0.4	1.0*	2.5	2.6	2.5	2.6	2.6	3.1^	2.2*	2.0*	24.2	36.9	37.1	36.7	37.6	37.6	39.2	34.6	33.4
0.6	0.4	1.5*	2.6	2.6	2.5	2.7^	2.7^	3.0^	2.3*	2.1*	25.7	33.5	33.6	33.3	34.1	34.1	35.3	31.6	30.4
0	0.5	0.4*	2.7^	2.8^	2.7^	2.9^	2.9^	3.6^	2.3*	2.1*	20.9	48.5	48.7	48.3	49.2	49.2	54.3	45.8	44.2
0.8	0.5	1.6*	2.7^	2.7^	2.7^	2.8^	2.8^	3.1^	2.4	2.2*	25.7	33.2	33.2	33.1	33.9	33.9	34.8	31.2	29.9
0.6	0.6	0.5*	2.7^	2.6	2.6	2.8^	2.8^	3.4^	2.3*	2.1*	22.0	44.1	44.0	43.8	44.8	44.8	46.9	41.5	39.5
0.8	0.8	0.1*	3.0^	2.8^	3.0^	3.1^	3.1^	4.2^	2.5	2.1*	18.1	57.9	57.0	57.7	58.7	58.7	61.9	54.9	51.9

Table 1.64. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=5, n_k~constant, p_C=0.1, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.2*	3.0^	2.9^	3.0^	3.1^	3.1^	3.1^	2.3*	2.2*	87.8	90.7	90.4	90.6	90.8	90.8	90.7	88.2	88.0
0.2	0	2.5	3.0^	2.9^	3.0^	3.1^	3.1^	3.1^	2.3*	2.2*	87.5	89.5	89.2	89.5	89.7	89.7	89.4	87.0	86.7
0.4	0	2.9^	3.1^	3.0^	3.0^	3.1^	3.1^	3.0^	2.4	2.3*	87.1	88.1	87.8	88.1	88.3	88.3	87.8	85.5	85.1
0.6	0	3.0^	2.7^	2.7^	2.7^	2.8^	2.8^	2.5	2.1*	2.1*	85.8	85.5	85.1	85.5	85.7	85.7	84.6	82.6	82.1
0.8	0	3.7^	2.8^	2.7^	2.8^	2.9^	2.9^	2.6	2.2*	2.1*	84.7	82.5	82.0	82.4	82.6	82.6	81.0	79.2	78.6
0	0.2	1.7*	3.1^	3.0^	3.1^	3.2^	3.2^	3.6^	2.4	2.3*	89.6	93.9	93.7	93.9	94.0	94.0	94.4	92.0	91.8
0.2	0.2	1.9*	3.2^	3.0^	3.1^	3.2^	3.2^	3.4^	2.4	2.3*	88.9	92.7	92.4	92.6	92.7	92.7	93.0	90.7	90.5
0.4	0.2	2.1*	3.0^	2.9^	3.0^	3.1^	3.1^	3.2^	2.3*	2.2*	88.4	91.4	91.1	91.4	91.5	91.5	91.6	89.1	88.8
0.6	0.2	2.6	3.1^	3.0^	3.1^	3.1^	3.1^	3.1^	2.4	2.2*	87.1	89.0	88.6	89.0	89.1	89.1	88.9	86.4	85.9
0.8	0.2	2.9^	2.8^	2.7^	2.8^	2.8^	2.8^	2.7^	2.2*	2.0*	86.1	86.1	85.6	86.0	86.2	86.2	85.3	83.3	82.6
0.4	0.4	1.3*	3.0^	2.9^	3.0^	3.0^	3.0^	3.5^	2.2*	2.1*	90.3	95.4	95.1	95.4	95.5	95.5	95.9	93.8	93.5
0.6	0.4	1.6*	3.0^	2.8^	3.0^	3.0^	3.0^	3.2^	2.2*	2.1*	89.1	93.3	93.0	93.3	93.4	93.4	93.6	91.4	91.0
0	0.5	0.8*	3.1^	3.0^	3.1^	3.2^	3.2^	4.1^	2.3*	2.1*	92.3	98.1	97.9	98.1	98.1	98.1	98.7	97.3	97.0
0.8	0.5	1.5*	2.8^	2.6	2.8^	2.8^	2.8^	3.0^	2.1*	1.9*	89.3	93.5	93.0	93.4	93.6	93.6	93.7	91.4	90.9
0.6	0.6	0.7*	3.1^	2.8^	3.1^	3.1^	3.1^	3.9^	2.2*	2.0*	92.1	97.7	97.5	97.7	97.7	97.7	98.1	96.8	96.5
0.8	0.8	0.2*	3.1^	2.7^	3.1^	3.2^	3.2^	4.6^	2.2*	1.9*	95.1	99.6	99.6	99.6	99.6	99.6	99.7	99.4	99.3

Table 1.65. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=5, n_k~constant, p_C=0.2, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.4	2.8^	2.8^	2.8^	2.8^	2.8^	2.9^	2.4	2.4	68.4	71.1	71.2	71.1	71.4	71.4	71.7	68.8	68.6
0.2	0	2.8^	2.8^	2.8^	2.8^	2.9^	2.9^	2.9^	2.5	2.5	68.3	68.8	68.9	68.7	69.1	69.1	68.8	66.4	66.0
0.4	0	3.1^	2.7^	2.7^	2.7^	2.8^	2.8^	2.7^	2.4	2.4	67.8	66.1	66.2	66.1	66.4	66.4	65.7	63.8	63.3
0.6	0	3.6^	2.6	2.6	2.6	2.7^	2.7^	2.6	2.3*	2.2*	67.0	62.5	62.5	62.4	62.7	62.7	61.8	60.1	59.4
0.8	0	4.4^	2.7^	2.7^	2.7^	2.7^	2.7^	2.6	2.4	2.3*	66.2	58.5	58.5	58.4	58.8	58.8	57.3	56.2	55.5
0	0.2	1.5*	2.8^	2.8^	2.8^	2.8^	2.8^	3.1^	2.4	2.3*	70.1	79.0	79.0	78.9	79.2	79.2	80.1	76.8	76.5
0.2	0.2	1.8*	2.9^	2.9^	2.9^	3.0^	3.0^	3.1^	2.5	2.4	69.7	75.9	75.9	75.9	76.1	76.1	76.6	73.8	73.4
0.4	0.2	2.2*	2.8^	2.8^	2.8^	2.8^	2.8^	2.9^	2.4	2.3*	69.1	72.7	72.7	72.7	73.0	73.0	73.0	70.4	69.9
0.6	0.2	2.5	2.5	2.5	2.5	2.6	2.6	2.6	2.3*	2.2*	68.4	69.0	69.0	69.0	69.3	69.3	69.0	66.7	66.1
0.8	0.2	3.1^	2.6	2.6	2.6	2.6	2.6	2.5	2.2*	2.1*	67.2	64.2	64.1	64.1	64.5	64.5	63.6	61.7	61.0
0.4	0.4	1.2*	2.7^	2.7^	2.7^	2.8^	2.8^	3.0^	2.3*	2.2*	70.7	81.5	81.4	81.5	81.7	81.7	82.5	79.4	78.8
0.6	0.4	1.5*	2.6	2.6	2.6	2.7^	2.7^	2.9^	2.2*	2.1*	70.4	77.7	77.6	77.7	77.9	77.9	78.2	75.5	74.8
0	0.5	0.6*	3.0^	3.0^	3.0^	3.1^	3.1^	3.6^	2.5	2.3*	73.9	91.1	91.0	91.1	91.2	91.2	93.1	89.6	89.2
0.8	0.5	1.6*	2.7^	2.7^	2.7^	2.8^	2.8^	2.9^	2.3*	2.2*	69.7	77.0	76.7	76.9	77.2	77.2	77.6	74.7	73.8
0.6	0.6	0.6*	2.8^	2.8^	2.8^	2.9^	2.9^	3.3^	2.4	2.2*	73.0	88.9	88.7	88.9	89.1	89.1	89.8	87.1	86.4
0.8	0.8	0.1*	2.8^	2.6	2.7^	2.8^	2.8^	3.5^	2.2*	1.9*	76.4	96.4	96.2	96.4	96.5	96.5	96.9	95.5	95.1

Table 1.66. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=5, n_k~constant, p_C=0.3, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.7^	2.7^	2.7^	2.7^	2.7^	2.8^	2.4	2.4	59.0	60.7	60.9	60.6	61.0	61.0	61.1	59.2	59.0
0.2	0	2.9^	2.7^	2.7^	2.6	2.7^	2.7^	2.8^	2.5	2.4	58.5	57.2	57.5	57.2	57.5	57.5	57.4	55.9	55.6
0.4	0	3.3^	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.4	2.4	58.0	54.0	54.3	54.0	54.3	54.3	54.0	52.8	52.4
0.6	0	3.8^	2.5	2.5	2.5	2.5	2.5	2.5	2.4	2.3*	57.4	50.5	50.7	50.4	50.8	50.8	50.3	49.3	48.7
0.8	0	4.4^	2.5	2.6	2.5	2.6	2.6	2.5	2.4	2.3*	57.3	46.9	47.1	46.8	47.2	47.2	46.5	45.8	45.2
0	0.2	1.3*	2.7^	2.7^	2.7^	2.7^	2.7^	3.0^	2.5	2.4	59.9	70.0	70.2	70.0	70.3	70.3	71.1	68.8	68.4
0.2	0.2	1.8*	2.7^	2.7^	2.7^	2.8^	2.8^	2.9^	2.5	2.4	58.9	65.4	65.6	65.3	65.7	65.7	66.1	64.0	63.5
0.4	0.2	2.2*	2.6	2.6	2.6	2.7^	2.7^	2.8^	2.4	2.3*	58.9	61.8	62.0	61.7	62.1	62.1	62.3	60.5	60.0
0.6	0.2	2.8^	2.6	2.7^	2.6	2.7^	2.7^	2.8^	2.5	2.4	58.2	57.6	57.8	57.6	57.9	57.9	57.8	56.3	55.7
0.8	0.2	3.4^	2.6	2.6	2.6	2.6	2.6	2.7^	2.4	2.3*	58.1	53.4	53.5	53.3	53.7	53.7	53.4	52.1	51.5
0.4	0.4	1.1*	2.7^	2.7^	2.7^	2.8^	2.8^	2.9^	2.5	2.4	59.7	72.2	72.3	72.1	72.5	72.5	73.1	70.9	70.2
0.6	0.4	1.6*	2.5	2.5	2.5	2.6	2.6	2.7^	2.3*	2.2*	59.3	67.2	67.3	67.2	67.5	67.5	67.8	66.0	65.2
0	0.5	0.4*	2.7^	2.7^	2.7^	2.7^	2.7^	3.1^	2.4	2.3*	61.2	85.2	85.2	85.2	85.4	85.4	87.7	84.2	83.5
0.8	0.5	1.5*	2.6	2.6	2.5	2.6	2.6	2.7^	2.4	2.2*	59.2	67.2	67.2	67.1	67.5	67.5	67.8	65.8	64.9
0.6	0.6	0.5*	2.6	2.6	2.6	2.7^	2.7^	3.0^	2.4	2.2*	60.8	80.9	80.8	80.8	81.1	81.1	81.9	79.7	78.8
0.8	0.8	0.1*	2.8^	2.6	2.7^	2.8^	2.8^	3.3^	2.4	2.2*	63.0	92.5	92.2	92.4	92.6	92.6	93.2	91.7	91.0

Table 1.67. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k=5, n_k~constant, p_C=0.1, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	3.0^	2.9^	3.0^	3.0^	3.0^	3.0^	2.5	2.5
0.2	0	2.6	2.9^	2.9^	2.9^	3.0^	3.0^	3.0^	2.4	2.3*
0.4	0	2.8^	2.8^	2.7^	2.8^	2.8^	2.8^	2.7^	2.4	2.3*
0.6	0	3.3^	2.8^	2.7^	2.8^	2.8^	2.8^	2.7^	2.4	2.3*
0.8	0	3.8^	2.6	2.6	2.6	2.7^	2.7^	2.4	2.2*	2.1*
0	0.2	1.7*	2.9^	2.8^	2.9^	2.9^	2.9^	3.1^	2.4	2.3*
0.2	0.2	1.8*	2.7^	2.6	2.7^	2.8^	2.8^	2.9^	2.3*	2.2*
0.4	0.2	2.1*	2.8^	2.7^	2.8^	2.8^	2.8^	2.9^	2.3*	2.2*
0.6	0.2	2.5	2.8^	2.8^	2.8^	2.9^	2.9^	2.8^	2.3*	2.2*
0.8	0.2	2.9^	2.6	2.5	2.6	2.6	2.6	2.5	2.1*	2.0*
0.4	0.4	1.3*	3.0^	2.8^	3.0^	3.0^	3.0^	3.2^	2.3*	2.2*
0.6	0.4	1.6*	2.8^	2.6	2.8^	2.8^	2.8^	2.9^	2.3*	2.2*
0	0.5	0.8*	3.0^	2.8^	3.0^	3.0^	3.0^	3.6^	2.3*	2.2*
0.8	0.5	1.8*	2.8^	2.7^	2.8^	2.8^	2.8^	3.0^	2.3*	2.2*
0.6	0.6	0.9*	3.2^	3.0^	3.2^	3.2^	3.2^	3.7^	2.5	2.4
0.8	0.8	0.3*	2.9^	2.5	2.9^	2.9^	2.9^	3.8^	2.2*	2.0*

Power (%)

FM	*MOM*	*ICC*	*O*	*GEE Ind*	*GEE Exc*	*GEE Un*	*AFM*	*AFM CC*
99.4	99.6	99.5	99.6	99.6	99.6	99.6	99.4	99.4
99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.2	99.2
99.3	99.3	99.3	99.3	99.3	99.3	99.2	99.1	99.1
99.0	98.9	98.9	98.9	98.9	98.9	98.8	98.6	98.6
98.7	98.3	98.2	98.3	98.3	98.3	98.1	97.9	97.9
99.6	99.8	99.8	99.8	99.8	99.8	99.9	99.8	99.8
99.6	99.8	99.8	99.8	99.8	99.8	99.8	99.7	99.7
99.5	99.6	99.6	99.6	99.6	99.6	99.7	99.5	99.5
99.4	99.5	99.5	99.5	99.5	99.5	99.5	99.3	99.3
99.2	99.1	99.0	99.1	99.1	99.1	99.0	98.8	98.8
99.8	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
99.7	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8
99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
99.6	99.8	99.8	99.8	99.8	99.8	99.8	99.7	99.7
99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.68. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k=5, n_k~constant, p_C=0.2, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.5	2.8^	2.9^	2.8^	2.9^	2.9^	2.9^	2.5	2.5	93.6	94.3	94.3	94.3	94.3	94.3	94.3	93.7	93.6
0.2	0	2.8^	2.7^	2.8^	2.7^	2.8^	2.8^	2.7^	2.5	2.4	93.0	92.9	92.9	92.9	93.0	93.0	92.9	92.3	92.2
0.4	0	3.2^	2.7^	2.7^	2.7^	2.7^	2.7^	2.7^	2.4	2.4	92.4	91.5	91.5	91.5	91.5	91.5	91.3	90.7	90.5
0.6	0	3.8^	2.7^	2.7^	2.7^	2.7^	2.7^	2.7^	2.5	2.5	91.8	89.5	89.5	89.5	89.5	89.5	89.1	88.7	88.5
0.8	0	4.5^	2.7^	2.7^	2.7^	2.7^	2.7^	2.6	2.5	2.4	90.7	86.4	86.5	86.4	86.5	86.5	85.9	85.5	85.2
0	0.2	1.5*	2.7^	2.7^	2.7^	2.7^	2.7^	2.9^	2.4	2.3*	95.3	97.3	97.3	97.3	97.3	97.3	97.5	97.0	97.0
0.2	0.2	1.8*	2.7^	2.7^	2.7^	2.7^	2.7^	2.8^	2.4	2.4	94.7	96.3	96.3	96.3	96.4	96.4	96.4	95.9	95.8
0.4	0.2	2.2*	2.7^	2.7^	2.7^	2.7^	2.7^	2.7^	2.4	2.3*	94.0	95.0	95.0	95.0	95.1	95.1	95.1	94.5	94.4
0.6	0.2	2.7^	2.7^	2.7^	2.7^	2.7^	2.7^	2.7^	2.4	2.3*	93.1	93.1	93.1	93.1	93.1	93.1	93.0	92.5	92.3
0.8	0.2	3.2^	2.5	2.5	2.5	2.5	2.5	2.5	2.3*	2.2*	92.2	90.6	90.6	90.6	90.7	90.7	90.5	89.9	89.6
0.4	0.4	1.1*	2.6	2.5	2.6	2.6	2.6	2.7^	2.3*	2.2*	95.7	97.9	97.9	97.9	98.0	98.0	98.1	97.7	97.6
0.6	0.4	1.6*	2.7^	2.6	2.7^	2.7^	2.7^	2.8^	2.4	2.3*	95.0	96.8	96.8	96.8	96.9	96.9	96.9	96.5	96.4
0	0.5	0.5*	2.6	2.6	2.6	2.6	2.6	3.0^	2.3*	2.2*	97.7	99.6	99.6	99.6	99.6	99.6	99.8	99.5	99.5
0.8	0.5	1.5*	2.5	2.5	2.5	2.6	2.6	2.7^	2.3*	2.2*	94.9	96.9	96.8	96.9	96.9	96.9	96.9	96.5	96.3
0.6	0.6	0.6*	2.7^	2.7^	2.7^	2.7^	2.7^	3.0^	2.4	2.3*	97.4	99.5	99.5	99.5	99.5	99.5	99.5	99.4	99.3
0.8	0.8	0.2*	2.7^	2.5	2.7^	2.7^	2.7^	3.2^	2.3*	2.1*	99.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9

Table 1.69. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k=5, n_k~constant, p_C=0.3, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.6	2.8^	2.8^	2.7^	2.8^	2.8^	2.8^	2.6	2.6	87.3	87.8	87.9	87.8	87.9	87.9	87.9	87.4	87.3
0.2	0	2.9^	2.6	2.6	2.6	2.6	2.6	2.6	2.4	2.4	86.5	85.6	85.7	85.6	85.7	85.7	85.5	85.1	84.9
0.4	0	3.2^	2.4	2.5	2.4	2.5	2.5	2.5	2.3*	2.3*	85.5	82.8	82.9	82.8	82.9	82.9	82.7	82.2	82.0
0.6	0	3.9^	2.6	2.7^	2.6	2.7^	2.7^	2.6	2.5	2.5	84.1	79.7	79.8	79.6	79.8	79.8	79.4	79.0	78.8
0.8	0	4.5^	2.6	2.6	2.6	2.6	2.6	2.5	2.5	2.4	83.3	76.0	76.1	76.0	76.2	76.2	75.6	75.4	75.0
0	0.2	1.5*	2.7^	2.8^	2.7^	2.8^	2.8^	2.9^	2.6	2.6	89.6	93.6	93.7	93.6	93.7	93.7	93.8	93.3	93.2
0.2	0.2	1.8*	2.7^	2.7^	2.7^	2.7^	2.7^	2.8^	2.5	2.5	88.2	91.0	91.1	91.0	91.1	91.1	91.2	90.7	90.5
0.4	0.2	2.2*	2.5	2.5	2.5	2.6	2.6	2.6	2.4	2.3*	87.7	89.1	89.1	89.1	89.1	89.1	89.1	88.6	88.4
0.6	0.2	2.7^	2.5	2.5	2.5	2.6	2.6	2.6	2.4	2.4	86.6	86.1	86.1	86.1	86.1	86.1	86.0	85.5	85.2
0.8	0.2	3.3^	2.5	2.5	2.5	2.5	2.5	2.5	2.4	2.3*	85.1	81.9	82.0	81.9	82.0	82.0	81.7	81.4	81.0
0.4	0.4	1.2*	2.7^	2.7^	2.7^	2.7^	2.7^	2.8^	2.6	2.5	90.2	94.8	94.8	94.8	94.9	94.9	95.1	94.6	94.4
0.6	0.4	1.5*	2.5	2.5	2.5	2.5	2.5	2.6	2.4	2.3*	89.3	92.6	92.6	92.6	92.6	92.6	92.7	92.2	92.0
0	0.5	0.3*	2.5	2.5	2.5	2.6	2.6	2.9^	2.3*	2.2*	93.6	98.8	98.8	98.8	98.8	98.8	99.1	98.7	98.6
0.8	0.5	1.7*	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.5	2.4	89.0	92.3	92.3	92.3	92.4	92.4	92.4	91.9	91.6
0.6	0.6	0.6*	2.6	2.6	2.6	2.6	2.6	2.8^	2.5	2.4	92.8	98.1	98.1	98.1	98.1	98.1	98.2	98.0	97.8
0.8	0.8	0.2*	2.8^	2.7^	2.8^	2.8^	2.8^	3.1^	2.6	2.4	96.0	99.7	99.7	99.7	99.8	99.8	99.8	99.7	99.7

Table 1.70. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k=5, n_k~constant, p_C=0.05, δ_0=-0.02

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.3*	3.4^	3.2^	3.4^	3.4^	3.4^	3.5^	2.5	2.4	67.7	73.8	73.1	73.7	73.9	73.9	74.0	68.6	68.2
0.2	0	2.5	3.4^	3.2^	3.4^	3.4^	3.4^	3.3^	2.4	2.4	67.6	72.5	71.9	72.5	72.6	72.6	72.3	67.3	66.9
0.4	0	2.6	3.1^	3.0^	3.1^	3.1^	3.1^	3.0^	2.3*	2.2*	67.3	70.6	69.8	70.5	70.7	70.7	69.9	65.3	64.8
0.6	0	2.9^	2.9^	2.8^	2.9^	3.0^	3.0^	2.7^	2.2*	2.1*	66.6	67.5	66.8	67.5	67.6	67.6	66.2	62.1	61.4
0.8	0	3.4^	2.8^	2.7^	2.8^	2.9^	2.9^	2.5	2.2*	2.0*	66.1	63.7	62.8	63.6	63.8	63.8	61.3	58.0	57.3
0	0.2	1.9*	3.4^	3.3^	3.4^	3.5^	3.5^	3.8^	2.4	2.4	68.5	77.6	76.9	77.5	77.7	77.7	78.5	72.5	72.1
0.2	0.2	1.9*	3.1^	3.0^	3.1^	3.2^	3.2^	3.3^	2.2*	2.2*	68.4	76.4	75.8	76.4	76.6	76.6	77.0	71.1	70.7
0.4	0.2	2.2*	3.2^	3.0^	3.2^	3.2^	3.2^	3.2^	2.3*	2.2*	67.2	73.6	72.9	73.6	73.8	73.8	73.6	68.2	67.6
0.6	0.2	2.4	3.0^	2.9^	3.0^	3.0^	3.0^	2.9^	2.2*	2.1*	66.9	70.8	70.0	70.8	71.0	71.0	70.2	65.5	64.7
0.8	0.2	2.9^	2.8^	2.7^	2.8^	2.8^	2.8^	2.6	2.0*	1.9*	66.1	66.4	65.4	66.4	66.5	66.5	64.9	61.0	60.1
0.4	0.4	1.5*	3.3^	3.2^	3.3^	3.3^	3.3^	3.7^	2.3*	2.2*	69.3	80.2	79.5	80.2	80.3	80.3	81.2	75.3	74.7
0.6	0.4	1.7*	3.0^	2.8^	3.0^	3.0^	3.0^	3.2^	2.1*	2.0*	68.3	76.9	76.0	76.9	77.0	77.0	77.3	71.6	70.7
0	0.5	1.0*	3.2^	3.0^	3.2^	3.2^	3.2^	3.9^	2.1*	2.0*	70.1	85.3	84.7	85.3	85.4	85.4	87.9	80.8	80.1
0.8	0.5	1.8*	3.1^	2.9^	3.1^	3.1^	3.1^	3.2^	2.1*	2.0*	68.4	76.0	74.9	76.0	76.2	76.2	76.1	70.7	69.7
0.6	0.6	1.0*	3.2^	3.0^	3.2^	3.2^	3.2^	3.8^	2.1*	2.0*	70.5	85.8	85.1	85.8	85.9	85.9	87.2	81.4	80.5
0.8	0.8	0.3*	3.3^	2.9^	3.2^	3.3^	3.3^	4.4^	2.1*	1.8*	73.3	94.1	93.4	94.1	94.1	94.1	95.2	91.1	90.1

Table 1.71. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=5, n_k~constant, p_C=0.05, δ_0=-0.02

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.2*	3.7^	3.6^	3.7^	3.8^	3.8^	3.9^	2.4	2.3*	37.7	47.9	46.8	47.8	48.1	48.1	48.3	39.2	38.8
0.2	0	2.2*	3.5^	3.3^	3.5^	3.5^	3.5^	3.5^	2.2*	2.2*	37.8	46.6	45.6	46.5	46.9	46.9	46.6	37.9	37.4
0.4	0	2.4	3.5^	3.3^	3.4^	3.5^	3.5^	3.3^	2.2*	2.1*	37.6	44.1	43.0	44.0	44.4	44.4	43.1	35.9	35.1
0.6	0	2.8^	3.3^	3.1^	3.2^	3.3^	3.3^	3.0^	2.1*	2.0*	38.3	41.9	40.7	41.9	42.2	42.2	40.0	33.9	33.0
0.8	0	3.1^	2.9^	2.7^	2.9^	3.0^	3.0^	[99]	1.9*	1.8*	39.0	38.9	37.5	38.8	39.2	39.2	[99]	31.1	30.0
0	0.2	1.6*	3.6^	3.5^	3.6^	3.7^	3.7^	4.2^	2.2*	2.1*	37.3	51.1	50.0	51.1	51.5	51.5	53.3	42.2	41.7
0.2	0.2	1.7*	3.5^	3.3^	3.5^	3.6^	3.6^	3.9^	2.2*	2.1*	37.4	49.9	48.8	49.8	50.2	50.2	51.1	40.8	40.1
0.4	0.2	2.0*	3.6^	3.3^	3.6^	3.7^	3.7^	3.7^	2.2*	2.0*	37.7	47.7	46.5	47.6	47.9	47.9	47.9	38.9	38.1
0.6	0.2	2.3*	3.4^	3.1^	3.3^	3.4^	3.4^	3.3^	2.1*	2.0*	37.9	44.8	43.5	44.7	45.0	45.0	43.9	36.3	35.3
0.8	0.2	2.7^	3.0^	2.8^	3.0^	3.0^	3.0^	[99]	1.8*	1.7*	38.2	40.8	39.3	40.7	41.1	41.1	[99]	32.7	31.5
0.4	0.4	1.3*	3.4^	3.1^	3.4^	3.4^	3.4^	3.9^	1.9*	1.8*	36.8	53.2	51.9	53.2	53.6	53.6	55.5	43.7	42.5
0.6	0.4	1.7*	3.3^	3.1^	3.3^	3.4^	3.4^	3.7^	2.0*	1.9*	37.1	49.8	48.3	49.8	50.1	50.1	50.7	40.6	39.3
0	0.5	0.9*	3.6^	3.4^	3.6^	3.7^	3.7^	4.8^	2.0*	1.9*	35.7	58.9	57.5	58.8	59.2	59.2	64.7	49.0	47.7
0.8	0.5	1.8*	3.2^	3.0^	3.2^	3.3^	3.3^	[99]	2.0*	1.8*	36.9	48.5	46.6	48.4	48.7	48.7	[99]	39.0	37.4
0.6	0.6	0.8*	3.4^	3.1^	3.4^	3.5^	3.5^	4.4^	1.9*	1.7*	35.5	59.5	57.6	59.4	59.8	59.8	63.0	49.0	47.2
0.8	0.8	0.3*	3.6^	3.1^	3.6^	3.7^	3.7^	5.5^	1.8*	1.5*	33.2	71.6	68.8	71.6	71.9	71.9	76.6	60.6	57.6

Table 1.72. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=5, n_k~constant, p_C=0.2, δ_0=-0.02

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.3*	2.8^	2.8^	2.8^	2.9^	2.9^	3.0^	2.4	2.4	16.1	18.4	18.4	18.3	18.5	18.5	18.8	16.4	16.2
0.2	0	2.8^	2.9^	2.9^	2.9^	2.9^	2.9^	2.9^	2.4	2.4	17.1	17.6	17.6	17.5	17.7	17.7	17.8	15.8	15.5
0.4	0	3.1^	2.7^	2.7^	2.7^	2.7^	2.7^	2.7^	2.3*	2.3*	17.9	16.4	16.5	16.4	16.6	16.6	16.3	14.8	14.5
0.6	0	3.6^	2.6	2.6	2.6	2.6	2.6	2.5	2.2*	2.2*	18.6	15.3	15.3	15.2	15.4	15.4	15.0	13.9	13.5
0.8	0	4.5^	2.8^	2.8^	2.7^	2.8^	2.8^	2.7^	2.4	2.3*	19.7	14.1	14.1	14.1	14.3	14.3	13.7	12.8	12.4
0	0.2	1.6*	3.0^	3.0^	3.0^	3.1^	3.1^	3.3^	2.5	2.4	13.7	20.9	20.9	20.8	21.1	21.1	22.2	18.6	18.3
0.2	0.2	1.8*	2.9^	2.9^	2.9^	2.9^	2.9^	3.1^	2.4	2.4	14.9	19.9	20.0	19.9	20.2	20.2	20.7	17.9	17.5
0.4	0.2	2.2*	2.8^	2.8^	2.8^	2.9^	2.9^	2.9^	2.4	2.3*	15.7	18.5	18.5	18.4	18.6	18.6	18.8	16.5	16.1
0.6	0.2	2.6	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.3*	2.1*	16.8	17.1	17.1	17.0	17.3	17.3	17.0	15.3	14.9
0.8	0.2	3.4^	2.7^	2.7^	2.6	2.7^	2.7^	2.6	2.3*	2.2*	18.4	15.8	15.8	15.8	16.0	16.0	15.6	14.2	13.7
0.4	0.4	1.2*	2.7^	2.7^	2.7^	2.8^	2.8^	3.0^	2.3*	2.1*	13.3	22.2	22.2	22.1	22.4	22.4	23.4	19.7	19.1
0.6	0.4	1.6*	2.7^	2.7^	2.7^	2.8^	2.8^	2.9^	2.3*	2.2*	14.4	19.9	19.9	19.9	20.2	20.2	20.5	17.9	17.3
0	0.5	0.4*	2.8^	2.8^	2.8^	2.8^	2.8^	3.4^	2.2*	2.0*	9.9	28.1	28.0	28.0	28.3	28.3	31.8	24.9	24.0
0.8	0.5	1.6*	2.6	2.6	2.6	2.7^	2.7^	2.8^	2.2*	2.0*	14.5	19.5	19.4	19.5	19.7	19.7	20.1	17.5	16.8
0.6	0.6	0.7*	2.8^	2.8^	2.8^	2.9^	2.9^	3.2^	2.3*	2.1*	10.5	25.8	25.7	25.7	26.0	26.0	27.4	22.7	21.8
0.8	0.8	0.1*	2.7^	2.6	2.7^	2.7^	2.7^	3.3^	2.0*	1.8*	6.8	33.8	33.4	33.7	34.0	34.0	36.6	29.8	28.0

Table 1.73. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=5, n_k~constant, p_C=0.2, δ_0=-0.1

Correlation Structure		Type I error rate (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.4	2.7^	2.7^	2.7^	2.7^	2.7^	2.9^	2.4	2.4
0.2	0	2.7^	2.6	2.6	2.6	2.7^	2.7^	2.7^	2.4	2.3*
0.4	0	3.1^	2.7^	2.7^	2.7^	2.7^	2.7^	2.7^	2.4	2.4
0.6	0	3.3^	2.4	2.5	2.4	2.5	2.5	2.4	2.2*	2.1*
0.8	0	4.1^	2.6	2.6	2.5	2.6	2.6	2.5	2.4	2.3*
0	0.2	1.6*	2.9^	2.8^	2.8^	2.9^	2.9^	3.2^	2.5	2.4
0.2	0.2	1.9*	2.8^	2.8^	2.8^	2.9^	2.9^	3.1^	2.5	2.5
0.4	0.2	2.2*	2.8^	2.8^	2.8^	2.9^	2.9^	2.9^	2.5	2.4
0.6	0.2	2.4	2.5	2.5	2.5	2.6	2.6	2.6	2.3*	2.2*
0.8	0.2	3.0^	2.5	2.5	2.5	2.5	2.5	2.5	2.3*	2.2*
0.4	0.4	1.2*	2.7^	2.6	2.7^	2.7^	2.7^	3.0^	2.4	2.3*
0.6	0.4	1.5*	2.6	2.5	2.6	2.6	2.6	2.8^	2.3*	2.2*
0	0.5	0.6*	2.8^	2.7^	2.8^	2.9^	2.9^	3.5^	2.4	2.3*
0.8	0.5	1.5*	2.6	2.5	2.6	2.6	2.6	2.8^	2.3*	2.2*
0.6	0.6	0.6*	2.8^	2.6	2.8^	2.8^	2.8^	3.3^	2.4	2.3*
0.8	0.8	0.2*	2.8^	2.4	2.8^	2.9^	2.9^	3.8^	2.4	2.2*

Power (%)

FM	*MOM*	*ICC*	*O*	*GEE Ind*	*GEE Exc*	*GEE Un*	*AFM*	*AFM CC*
99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8
99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.7	99.7
99.7	99.6	99.6	99.6	99.6	99.6	99.6	99.5	99.5
99.5	99.3	99.3	99.3	99.3	99.3	99.3	99.2	99.2
99.5	99.1	99.1	99.1	99.1	99.1	99.0	98.9	98.9
99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
99.8	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.8
99.7	99.8	99.8	99.8	99.8	99.8	99.8	99.7	99.7
99.6	99.6	99.5	99.5	99.6	99.6	99.5	99.5	99.4
99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
99.9	100.0	99.9	100.0	100.0	100.0	100.0	99.9	99.9
100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.74. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=5, n_k~constant, p_C=0.3, δ_0=-0.1

Correlation Structure		***Type I error rate (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.6	2.6	2.6	2.6	2.6	2.8^	2.5	2.4
0.2	0	2.8^	2.6	2.6	2.6	2.6	2.6	2.7^	2.5	2.4
0.4	0	3.3^	2.5	2.6	2.5	2.6	2.6	2.6	2.5	2.4
0.6	0	3.8^	2.5	2.6	2.5	2.6	2.6	2.5	2.4	2.4
0.8	0	4.5^	2.6	2.7^	2.6	2.7^	2.7^	2.6	2.5	2.4
0	0.2	1.4*	2.5	2.6	2.5	2.6	2.6	2.8^	2.5	2.4
0.2	0.2	1.8*	2.7^	2.7^	2.6	2.7^	2.7^	2.9^	2.6	2.5
0.4	0.2	2.2*	2.6	2.6	2.6	2.6	2.6	2.7^	2.5	2.4
0.6	0.2	2.8^	2.7^	2.7^	2.6	2.7^	2.7^	2.8^	2.6	2.5
0.8	0.2	3.2^	2.4	2.5	2.4	2.5	2.5	2.4	2.4	2.3*
0.4	0.4	1.1*	2.6	2.6	2.6	2.7^	2.7^	2.8^	2.5	2.4
0.6	0.4	1.5*	2.5	2.5	2.5	2.6	2.6	2.7^	2.4	2.3*
0	0.5	0.4*	2.6	2.5	2.6	2.6	2.6	3.0^	2.4	2.3*
0.8	0.5	1.6*	2.6	2.6	2.6	2.7^	2.7^	2.9^	2.6	2.4
0.6	0.6	0.6*	2.6	2.5	2.6	2.6	2.6	2.9^	2.5	2.3*
0.8	0.8	0.1*	2.7^	2.4	2.7^	2.7^	2.7^	3.4^	2.5	2.3*

Power (%)

FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
99.1	99.2	99.2	99.2	99.2	99.2	99.2	99.1	99.1
98.9	98.8	98.8	98.8	98.8	98.8	98.8	98.7	98.7
98.7	98.3	98.4	98.3	98.4	98.4	98.2	98.2	98.2
98.3	97.5	97.5	97.5	97.5	97.5	97.4	97.4	97.3
98.0	96.4	96.5	96.4	96.5	96.5	96.2	96.3	96.1
99.6	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8
99.4	99.6	99.6	99.6	99.6	99.6	99.6	99.6	99.6
99.2	99.4	99.4	99.4	99.4	99.4	99.4	99.3	99.3
99.0	98.9	98.9	98.9	98.9	98.9	98.9	98.9	98.8
98.6	98.2	98.3	98.2	98.3	98.3	98.2	98.2	98.1
99.7	99.9	99.9	99.9	99.9	99.9	99.9	99.9	99.9
99.5	99.8	99.8	99.8	99.8	99.8	99.8	99.7	99.7
99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
99.5	99.7	99.7	99.7	99.7	99.7	99.7	99.7	99.7
99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.75. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k=5, n_k~constant, p_C=0.5, δ_0=-0.1

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.3*	2.4	2.3*	2.4	2.4	2.5	2.6	2.5	98.4	98.3	98.3	98.3	98.3	98.3	98.3	98.4	98.3
0.2	0	3.0^	2.4	2.4	2.4	2.4	2.4	2.5	2.5	2.5	97.9	97.3	97.4	97.3	97.4	97.4	97.4	97.5	97.4
0.4	0	3.6^	2.3*	2.4	2.3*	2.3*	2.3*	2.4	2.4	2.4	97.7	96.4	96.5	96.4	96.5	96.5	96.5	96.6	96.5
0.6	0	4.2^	2.4	2.4	2.4	2.4	2.4	2.5	2.5	2.4	97.1	95.2	95.3	95.1	95.2	95.2	95.2	95.3	95.2
0.8	0	5.2^	2.6	2.6	2.6	2.6	2.6	2.7^	2.7^	2.6	96.5	93.2	93.4	93.2	93.3	93.3	93.3	93.5	93.3
0	0.2	1.3*	2.4	2.4	2.4	2.4	2.4	2.5	2.5	2.5	99.3	99.6	99.6	99.6	99.6	99.6	99.6	99.7	99.7
0.2	0.2	1.9*	2.4	2.4	2.4	2.5	2.5	2.6	2.6	2.5	98.9	99.2	99.2	99.2	99.2	99.2	99.2	99.2	99.2
0.4	0.2	2.4	2.4	2.5	2.4	2.5	2.5	2.6	2.6	2.5	98.6	98.6	98.6	98.6	98.6	98.6	98.7	98.7	98.7
0.6	0.2	2.9^	2.4	2.4	2.4	2.4	2.4	2.5	2.6	2.5	98.1	97.8	97.8	97.8	97.8	97.8	97.8	97.9	97.8
0.8	0.2	3.7^	2.3*	2.4	2.3*	2.4	2.4	2.5	2.5	2.4	97.6	96.5	96.5	96.5	96.5	96.5	96.5	96.6	96.5
0.4	0.4	1.1*	2.4	2.4	2.4	2.4	2.4	2.6	2.6	2.5	99.4	99.7	99.7	99.7	99.7	99.7	99.7	99.7	99.7
0.6	0.4	1.7*	2.4	2.4	2.4	2.4	2.4	2.5	2.6	2.4	99.0	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4
0	0.5	0.3*	2.4	2.4	2.4	2.5	2.5	2.6	2.6	2.5	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.8	0.5	1.8*	2.5	2.5	2.5	2.5	2.5	2.7^	2.7^	2.5	99.0	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4
0.6	0.6	0.6*	2.5	2.4	2.4	2.5	2.5	2.7^	2.7^	2.5	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.8	0.8	0.1*	2.6	2.3*	2.6	2.6	2.6	2.9^	2.9^	2.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.76. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k≤5, n_k~uniform, p_C=0.1, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.2*	3.2^	3.6^	3.7^	3.6^	3.6^	[99]	2.4	2.3*	45.4	49.8	47.1	47.8	53.9	53.9	[99]	47.3	46.5
0.2	0	2.6	3.2^	3.6^	3.7^	3.4^	3.4^	[99]	2.6	2.4	45.4	49.0	45.7	46.5	52.0	52.0	[99]	45.3	44.4
0.4	0	2.7^	3.0^	3.3^	3.3^	3.1^	3.1^	[97]	2.3*	2.1*	45.5	47.5	43.7	44.6	49.8	49.9	[96]	43.3	42.2
0.6	0	3.1^	3.0^	3.2^	3.3^	3.1^	3.2^	[91]	2.3*	2.1*	45.0	45.6	41.1	41.9	46.4	46.8	[90]	40.0	38.8
0.8	0	3.7^	2.9^	3.0^	3.0^	2.9^	3.0^	[78]	2.2*	2.0*	45.7	43.7	38.2	39.4	43.3	44.3	[76]	37.1	35.5
0	0.2	1.7*	3.2^	3.6^	3.7^	3.5^	3.5^	[99]	2.5	2.4	45.4	54.6	51.9	52.7	59.1	59.0	[99]	52.2	51.1
0.2	0.2	1.7*	3.0^	3.5^	3.6^	3.4^	3.3^	[99]	2.3*	2.1*	45.4	53.2	49.9	50.8	56.6	56.7	[99]	49.8	48.6
0.4	0.2	2.1*	3.0^	3.4^	3.5^	3.2^	3.2^	[97]	2.3*	2.1*	44.9	51.1	47.2	48.2	53.6	53.8	[97]	46.8	45.4
0.6	0.2	2.4	3.0^	3.1^	3.3^	3.1^	3.1^	[92]	2.2*	2.0*	45.0	49.2	44.6	45.8	50.3	50.9	[91]	43.6	42.1
0.8	0.2	3.0^	2.9^	2.9^	3.1^	2.9^	3.0^	[79]	2.1*	1.9*	45.0	46.8	41.0	42.3	46.2	47.3	[78]	39.7	38.0
0.4	0.4	1.4*	3.2^	3.5^	3.7^	3.4^	3.5^	[97]	2.4	2.1*	44.7	57.6	53.6	55.0	60.5	60.7	[97]	53.4	51.8
0.6	0.4	1.8*	3.2^	3.2^	3.4^	3.4^	3.3^	[92]	2.4	2.2*	45.2	55.8	50.5	52.2	57.0	57.8	[91]	49.7	47.9
0	0.5	0.8*	3.3^	3.5^	3.8^	3.5^	3.5^	[97]	2.4	2.2*	44.1	65.6	62.8	64.3	69.8	70.1	[96]	62.8	61.1
0.8	0.5	1.7*	2.9^	3.0^	3.3^	3.0^	3.1^	[81]	2.2*	1.9*	45.1	56.5	49.5	51.8	55.8	57.5	[80]	48.4	46.2
0.6	0.6	0.9*	3.2^	3.3^	3.6^	3.4^	3.4^	[93]	2.3*	2.0*	44.3	65.3	60.3	62.5	66.6	67.8	[92]	58.9	56.5
0.8	0.8	0.4*	3.4^	3.2^	3.9^	3.5^	3.6^	[85]	2.4	1.9*	43.3	78.6	71.9	75.6	76.6	80.4	[83]	69.1	65.9

Table 1.77. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k≤5, n_k~uniform, p_C=0.2, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.3*	2.8^	3.3^	3.1^	3.0^	3.0^	[99]	2.5	2.4	30.4	31.5	30.9	30.4	34.9	34.9	[99]	31.6	31.0
0.2	0	2.7^	2.7^	3.0^	2.9^	3.0^	3.0^	[99]	2.5	2.4	30.9	30.6	29.1	28.8	32.8	32.8	[99]	29.7	29.0
0.4	0	3.1^	2.8^	3.1^	3.0^	2.8^	2.9^	[99]	2.4	2.3*	31.5	29.4	27.9	27.6	30.7	30.9	[99]	28.0	27.1
0.6	0	3.4^	2.6	2.7^	2.7^	2.6	2.7^	[97]	2.2*	2.1*	32.0	27.9	25.7	25.4	28.6	28.9	[97]	25.9	25.0
0.8	0	4.0^	2.5	2.8^	2.7^	2.5	2.6	[91]	2.1*	2.0*	32.4	26.5	23.7	23.5	26.0	26.8	[91]	23.6	22.6
0	0.2	1.6*	3.0^	3.2^	3.1^	3.2^	3.2^	[99]	2.6	2.5	29.0	35.7	35.5	35.1	39.7	39.7	[99]	36.2	35.4
0.2	0.2	1.8*	2.8^	3.1^	3.1^	3.0^	3.0^	[99]	2.5	2.3*	29.7	34.4	33.5	33.2	37.1	37.2	[99]	33.7	32.8
0.4	0.2	2.1*	2.6	3.0^	3.0^	2.7^	2.8^	[99]	2.3*	2.2*	30.2	32.9	31.2	30.9	34.6	34.9	[99]	31.3	30.4
0.6	0.2	2.6	2.6	2.8^	2.7^	2.7^	2.8^	[98]	2.3*	2.2*	31.2	31.5	29.2	29.0	32.2	32.8	[97]	29.3	28.1
0.8	0.2	3.2^	2.5	2.6	2.6	2.7^	2.7^	[92]	2.3*	2.1*	32.2	29.9	26.9	26.8	29.5	30.5	[91]	26.8	25.6
0.4	0.4	1.3*	2.7^	3.0^	3.0^	3.0^	3.0^	[99]	2.5	2.3*	28.5	38.7	37.0	37.0	40.8	41.3	[99]	37.2	35.8
0.6	0.4	1.7*	2.7^	2.9^	2.9^	2.9^	2.8^	[98]	2.4	2.2*	29.5	36.4	34.1	34.1	37.2	38.2	[98]	33.8	32.4
0	0.5	0.6*	2.8^	3.2^	3.2^	3.1^	3.1^	[98]	2.5	2.3*	25.6	46.8	47.1	47.1	51.7	52.0	[97]	47.9	46.3
0.8	0.5	1.8*	2.7^	2.8^	2.8^	2.8^	2.9^	[93]	2.3*	2.1*	29.6	37.6	33.6	33.9	36.4	38.3	[92]	32.9	31.3
0.6	0.6	0.9*	2.9^	3.1^	3.1^	3.0^	3.0^	[98]	2.5	2.2*	27.0	45.8	42.9	43.3	46.3	48.0	[98]	42.2	40.4
0.8	0.8	0.3*	2.9^	2.8^	3.1^	3.0^	3.0^	[94]	2.5	2.2*	23.8	59.3	55.1	56.7	56.0	61.6	[94]	51.5	48.5

Table 1.78. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k≤5, n_k~uniform, p_C=0.3, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.6	2.9^	2.8^	2.8^	2.8^	3.1^	2.5	2.5	25.5	25.1	24.8	24.2	27.8	27.7	28.8	25.9	25.4
0.2	0	2.8^	2.5	2.7^	2.6	2.7^	2.7^	[99]	2.4	2.3*	26.4	24.2	23.7	23.2	26.2	26.3	[99]	24.6	24.0
0.4	0	3.3^	2.6	2.8^	2.7^	2.8^	2.8^	[99]	2.5	2.4	26.6	22.9	21.8	21.4	24.0	24.2	[99]	22.5	21.8
0.6	0	4.0^	2.5	2.8^	2.7^	2.7^	2.7^	[99]	2.4	2.3*	27.6	21.9	20.6	20.2	22.4	22.8	[99]	21.0	20.3
0.8	0	4.4^	2.5	2.6	2.5	2.6	2.6	[97]	2.4	2.3*	28.2	21.1	19.2	18.8	20.8	21.5	[97]	19.6	18.9
0	0.2	1.5*	2.6	2.8^	2.7^	2.9^	2.9^	3.4^	2.6	2.4	23.3	28.8	28.8	28.2	32.2	32.3	[99]	30.3	29.6
0.2	0.2	1.8*	2.5	2.7^	2.6	2.7^	2.7^	[99]	2.4	2.3*	24.5	28.0	27.4	26.8	30.1	30.3	[99]	28.4	27.6
0.4	0.2	2.4	2.8^	2.9^	2.8^	2.9^	2.9^	[99]	2.6	2.5	25.3	26.5	25.3	24.8	27.9	28.2	[99]	26.1	25.3
0.6	0.2	2.7^	2.5	2.6	2.5	2.5	2.5	[99]	2.3*	2.2*	26.0	25.2	23.3	22.9	25.6	26.2	[99]	24.0	23.1
0.8	0.2	3.4^	2.5	2.5	2.4	2.6	2.6	[97]	2.4	2.3*	27.3	24.1	21.9	21.5	23.6	24.6	[97]	22.2	21.3
0.4	0.4	1.3*	2.6	2.8^	2.7^	2.8^	2.8^	[99]	2.5	2.4	23.0	31.6	30.1	29.7	33.2	33.9	[99]	31.4	30.2
0.6	0.4	1.8*	2.6	2.7^	2.7^	2.7^	2.7^	[99]	2.4	2.2*	24.0	29.8	27.6	27.3	29.9	30.9	[99]	28.2	27.0
0	0.5	0.5*	2.6	2.8^	2.7^	2.9^	2.7^	[98]	2.5	2.3*	18.8	39.0	39.9	39.4	43.7	44.2	[98]	41.8	40.4
0.8	0.5	1.8*	2.5	2.6	2.5	2.6	2.6	[97]	2.4	2.2*	23.8	30.1	27.3	27.1	29.0	31.1	[97]	27.4	26.1
0.6	0.6	0.8*	2.7^	2.8^	2.8^	2.8^	2.8^	[99]	2.5	2.3*	20.4	37.0	35.6	35.4	37.5	39.5	[99]	35.4	33.7
0.8	0.8	0.3*	2.9^	2.9^	3.0^	2.8^	3.0^	[98]	2.7^	2.3*	17.6	50.4	47.3	48.1	46.6	52.7	[97]	44.4	41.9

Table 1.79. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, $n_k \leq 5$, n_k~uniform, $p_C=0.1$, $\delta_0=-0.05$

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.3*	2.9^	3.2^	3.2^	3.1^	3.1^	3.2^	2.5	2.4	85.3	85.6	81.9	82.3	88.1	88.1	88.2	86.1	85.8
0.2	0	2.6	2.8^	3.0^	3.1^	3.0^	3.0^	[99]	2.5	2.4	84.9	84.8	80.9	81.4	86.8	86.9	86.4	84.6	84.2
0.4	0	2.8^	2.8^	3.0^	3.0^	2.9^	2.8^	[99]	2.4	2.3*	84.2	83.8	78.9	79.4	85.1	85.2	[99]	82.5	82.0
0.6	0	3.1^	2.7^	3.0^	3.0^	2.8^	2.8^	[99]	2.3*	2.2*	83.4	82.2	76.5	77.0	82.6	83.0	[99]	80.0	79.4
0.8	0	3.7^	2.7^	2.7^	2.8^	2.7^	2.7^	[95]	2.2*	2.1*	82.6	80.3	73.3	74.0	79.6	80.4	[95]	76.5	75.7
0	0.2	1.8*	2.9^	3.2^	3.2^	3.1^	3.1^	3.5^	2.5	2.4	87.1	89.6	86.4	86.8	91.8	91.8	92.2	90.0	89.7
0.2	0.2	2.0*	2.9^	3.1^	3.2^	3.1^	3.1^	3.2^	2.5	2.4	86.0	88.2	84.7	85.2	90.1	90.2	90.1	88.2	87.8
0.4	0.2	2.1*	2.8^	3.0^	3.1^	2.9^	2.9^	[99]	2.3*	2.2*	85.7	87.3	83.2	83.8	88.6	88.8	[99]	86.5	86.1
0.6	0.2	2.4	2.6	2.8^	2.8^	2.7^	2.7^	[99]	2.2*	2.1*	84.1	85.3	80.1	80.8	85.8	86.2	[99]	83.4	82.7
0.8	0.2	3.1^	2.7^	2.8^	2.9^	2.7^	2.7^	[96]	2.3*	2.2*	83.6	83.8	77.2	77.9	82.9	83.9	[95]	80.2	79.4
0.4	0.4	1.5*	3.0^	3.1^	3.2^	3.1^	3.1^	[99]	2.5	2.4	87.5	91.9	88.7	89.3	93.0	93.2	[99]	91.3	90.8
0.6	0.4	1.8*	2.9^	2.9^	3.1^	2.9^	2.9^	[99]	2.3*	2.2*	86.4	90.4	86.2	86.9	90.7	91.2	[99]	88.7	88.1
0	0.5	0.8*	2.9^	3.0^	3.2^	3.0^	3.0^	[99]	2.4	2.2*	90.0	95.6	94.2	94.6	96.9	97.0	[99]	96.0	95.8
0.8	0.5	1.8*	2.8^	2.8^	3.0^	2.8^	2.8^	[96]	2.4	2.2*	86.4	91.2	86.5	87.4	90.4	91.5	[96]	88.3	87.6
0.6	0.6	0.9*	2.9^	2.9^	3.1^	3.0^	2.9^	[99]	2.3*	2.1*	89.1	95.6	93.2	93.8	95.8	96.3	[99]	94.5	94.1
0.8	0.8	0.4*	3.1^	3.0^	3.6^	3.1^	3.3^	[98]	2.5	2.2*	92.0	99.0	98.0	98.4	98.6	99.2	[97]	97.9	97.7

Table 1.80. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤5, n_k~uniform, p_C=0.2, $δ_0$=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.4	2.7^	2.9^	2.8^	2.8^	2.8^	2.9^	2.5	2.4	64.2	63.1	60.3	60.2	66.9	66.9	67.2	65.0	64.5
0.2	0	2.8^	2.6	2.8^	2.7^	2.8^	2.8^	2.8^	2.5	2.4	64.5	61.9	58.5	58.4	64.8	64.8	64.2	62.9	62.4
0.4	0	3.2^	2.6	2.9^	2.8^	2.7^	2.7^	[99]	2.5	2.4	63.9	60.0	55.6	55.5	61.7	62.0	60.9	59.8	59.1
0.6	0	3.7^	2.6	2.8^	2.7^	2.7^	2.7^	[99]	2.4	2.3*	63.0	58.1	52.7	52.6	58.0	58.7	[99]	56.1	55.4
0.8	0	4.3^	2.6	2.6	2.6	2.6	2.6	[99]	2.4	2.3*	63.0	55.8	49.8	49.7	54.5	55.8	[99]	52.5	51.8
0	0.2	1.6*	2.6	3.0^	3.0^	2.9^	2.9^	3.2^	2.5	2.5	65.8	69.7	67.8	67.7	74.3	74.2	75.2	72.4	72.0
0.2	0.2	1.9*	2.6	2.8^	2.8^	2.8^	2.8^	2.9^	2.5	2.4	65.4	68.2	65.2	65.2	71.4	71.6	71.6	69.5	68.9
0.4	0.2	2.4	2.8^	2.9^	2.9^	2.8^	2.9^	[99]	2.5	2.4	65.0	66.4	62.3	62.2	68.0	68.4	[99]	66.1	65.5
0.6	0.2	2.9^	2.7^	2.9^	2.9^	2.8^	2.8^	[99]	2.6	2.4	64.3	63.8	58.8	58.8	64.1	65.1	[99]	62.2	61.4
0.8	0.2	3.3^	2.5	2.7^	2.7^	2.5	2.6	[99]	2.3*	2.2*	63.8	61.2	54.8	54.8	59.8	61.4	[99]	57.8	56.9
0.4	0.4	1.3*	2.6	2.7^	2.7^	2.7^	2.7^	2.8^	2.4	2.3*	66.4	74.3	71.2	71.3	76.2	76.9	77.0	74.4	73.6
0.6	0.4	1.8*	2.6	2.8^	2.8^	2.6	2.7^	[99]	2.4	2.2*	65.6	71.8	67.1	67.3	72.1	73.4	[99]	70.2	69.3
0	0.5	0.6*	2.9^	3.0^	3.0^	2.9^	3.0^	[99]	2.6	2.5	68.7	83.7	83.5	83.6	87.3	87.9	[99]	86.2	85.7
0.8	0.5	1.8*	2.6	2.8^	2.8^	2.5	2.6	[99]	2.3*	2.1*	65.3	73.1	67.3	67.6	71.0	73.8	[99]	69.1	68.2
0.6	0.6	0.8*	2.7^	2.8^	2.9^	2.8^	2.8^	[99]	2.4	2.3*	67.5	82.5	79.2	79.5	82.4	84.3	[99]	80.7	79.9
0.8	0.8	0.3*	2.8^	2.7^	2.9^	2.8^	2.8^	[99]	2.5	2.2*	69.6	93.1	90.7	91.3	90.8	94.0	[99]	89.5	88.6

Table 1.81. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤5, n_k~uniform, p_C=0.3, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.4	2.6	2.8^	2.8^	2.7^	2.7^	2.8^	2.5	2.5	54.7	52.3	50.3	50.0	56.3	56.4	56.7	55.2	54.8
0.2	0	2.9^	2.6	2.7^	2.6	2.6	2.6	2.6	2.5	2.4	54.4	50.7	47.6	47.2	53.1	53.2	53.1	52.0	51.5
0.4	0	3.3^	2.5	2.6	2.5	2.6	2.6	2.6	2.4	2.3*	53.8	48.5	44.7	44.4	49.8	50.3	49.6	48.7	48.2
0.6	0	3.9^	2.5	2.6	2.5	2.6	2.6	[99]	2.5	2.4	53.8	46.8	42.4	42.1	46.7	47.5	[99]	45.8	45.2
0.8	0	4.6^	2.6	2.7^	2.6	2.5	2.6	[99]	2.4	2.3*	53.7	44.4	39.7	39.4	43.3	44.6	[99]	42.3	41.7
0	0.2	1.5*	2.6	2.7^	2.7^	2.8^	2.8^	3.0^	2.6	2.5	54.7	59.4	57.8	57.5	64.1	64.1	65.1	63.0	62.5
0.2	0.2	2.0*	2.7^	2.9^	2.8^	2.8^	2.8^	2.8^	2.6	2.5	54.6	57.2	54.5	54.2	60.3	60.4	60.5	59.2	58.6
0.4	0.2	2.3*	2.6	2.8^	2.7^	2.6	2.7^	2.7^	2.5	2.4	54.5	55.2	51.8	51.5	56.7	57.3	57.0	55.5	54.9
0.6	0.2	2.9^	2.5	2.7^	2.7^	2.6	2.6	2.6	2.5	2.4	54.3	53.0	48.5	48.3	52.8	54.0	53.5	51.7	51.0
0.8	0.2	3.5^	2.5	2.6	2.5	2.6	2.6	[99]	2.5	2.3*	53.8	50.1	44.9	44.7	48.4	50.3	[99]	47.4	46.6
0.4	0.4	1.3*	2.6	2.7^	2.6	2.6	2.6	2.8^	2.5	2.4	55.4	64.3	61.4	61.3	66.1	67.2	67.4	65.1	64.3
0.6	0.4	1.8*	2.6	2.7^	2.6	2.6	2.7^	2.7^	2.5	2.4	55.4	61.8	57.6	57.4	61.6	63.3	63.0	60.4	59.5
0	0.5	0.4*	2.5	2.8^	2.7^	2.7^	2.7^	[99]	2.6	2.5	56.3	75.9	76.2	76.1	80.6	81.3	[99]	80.0	79.3
0.8	0.5	1.9*	2.6	2.7^	2.6	2.6	2.7^	[99]	2.5	2.4	55.0	63.0	57.5	57.5	60.6	63.8	[99]	59.4	58.5
0.6	0.6	0.8*	2.6	2.7^	2.7^	2.7^	2.6	2.8^	2.5	2.3*	55.6	73.2	70.4	70.5	72.9	75.5	75.7	71.8	70.9
0.8	0.8	0.3*	2.8^	2.7^	2.8^	2.8^	2.8^	[99]	2.8^	2.5	56.0	87.1	84.2	84.6	83.2	88.5	[99]	82.3	81.1

Table 1.82. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k≤5, n_k~uniform, p_C=0.1, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.9^	3.1^	3.1^	3.0^	3.0^	3.1^	2.6	2.6	99.0	98.9	98.1	98.2	99.3	99.3	99.3	99.1	99.1
0.2	0	2.6	2.7^	2.9^	2.9^	2.8^	2.8^	2.7^	2.5	2.4	99.0	98.8	97.7	97.8	99.1	99.1	99.0	98.9	98.9
0.4	0	2.9^	2.7^	2.9^	2.9^	2.7^	2.8^	2.6	2.4	2.4	98.8	98.5	97.2	97.3	98.8	98.8	98.7	98.6	98.5
0.6	0	3.2^	2.6	2.7^	2.8^	2.7^	2.7^	[99]	2.4	2.3*	98.6	98.3	96.6	96.7	98.4	98.4	[99]	98.0	98.0
0.8	0	3.8^	2.6	2.7^	2.8^	2.6	2.6	[99]	2.3*	2.2*	98.2	97.8	95.3	95.5	97.4	97.7	[99]	97.0	96.9
0	0.2	1.8*	2.7^	3.0^	3.1^	2.9^	2.9^	3.1^	2.5	2.4	99.4	99.5	99.0	99.1	99.7	99.7	99.7	99.6	99.6
0.2	0.2	2.0*	2.8^	3.0^	3.1^	3.0^	3.0^	3.0^	2.6	2.5	99.3	99.4	98.8	98.9	99.6	99.6	99.6	99.5	99.5
0.4	0.2	2.2*	2.8^	2.8^	2.9^	2.7^	2.7^	2.7^	2.4	2.3*	99.0	99.1	98.3	98.4	99.3	99.3	99.3	99.1	99.1
0.6	0.2	2.7^	2.8^	2.8^	2.8^	2.8^	2.8^	[99]	2.4	2.3*	99.0	99.0	98.0	98.1	99.1	99.2	[99]	98.9	98.9
0.8	0.2	3.1^	2.6	2.6	2.6	2.6	2.6	[99]	2.3*	2.2*	98.6	98.7	96.9	97.1	98.4	98.7	[99]	98.1	98.0
0.4	0.4	1.5*	2.7^	2.9^	3.0^	2.9^	2.8^	2.9^	2.4	2.3*	99.5	99.7	99.4	99.4	99.8	99.8	99.8	99.7	99.7
0.6	0.4	1.7*	2.7^	2.7^	2.9^	2.7^	2.7^	[99]	2.3*	2.2*	99.3	99.6	99.1	99.1	99.6	99.7	[99]	99.5	99.5
0	0.5	0.8*	2.8^	2.8^	2.9^	2.8^	2.8^	[99]	2.4	2.3*	99.8	99.9	99.8	99.9	100.0	100.0	[99]	100.0	100.0
0.8	0.5	1.9*	2.6	2.7^	2.8^	2.7^	2.7^	[99]	2.4	2.3*	99.4	99.6	99.1	99.2	99.6	99.7	[99]	99.5	99.5
0.6	0.6	0.9*	2.7^	2.8^	2.9^	2.8^	2.8^	[99]	2.4	2.2*	99.7	99.9	99.8	99.8	99.9	100.0	[99]	99.9	99.9
0.8	0.8	0.4*	2.9^	2.7^	3.0^	2.9^	2.9^	[99]	2.5	2.3*	99.9	100.0	100.0	100.0	100.0	100.0	[99]	100.0	100.0

Table 1.83. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, $n_k \leq 5$, n_k~uniform, p_C=0.2, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.5	2.7^	2.8^	2.8^	2.8^	2.8^	2.8^	2.6	2.5	91.4	90.2	87.7	87.7	92.2	92.2	92.2	91.6	91.5
0.2	0	2.7^	2.6	2.7^	2.7^	2.6	2.6	2.6	2.5	2.4	90.8	88.9	86.0	86.0	90.6	90.6	90.4	90.0	89.8
0.4	0	3.2^	2.6	2.7^	2.7^	2.7^	2.6	2.5	2.5	2.4	90.4	88.0	84.2	84.2	88.9	89.2	88.7	88.3	88.1
0.6	0	3.6^	2.5	2.5	2.5	2.5	2.5	2.3*	2.3*	2.3*	89.4	86.4	81.6	81.6	86.4	86.9	86.4	85.6	85.4
0.8	0	4.4^	2.6	2.6	2.6	2.5	2.6	[99]	2.4	2.3*	88.2	84.3	78.5	78.5	83.2	84.3	[99]	82.3	81.9
0	0.2	1.6*	2.7^	2.9^	2.9^	2.7^	2.8^	2.9^	2.5	2.4	93.1	93.6	92.3	92.3	95.6	95.6	95.8	95.3	95.1
0.2	0.2	1.9*	2.6	2.8^	2.8^	2.7^	2.7^	2.7^	2.5	2.4	92.5	92.9	90.8	90.8	94.4	94.5	94.4	93.9	93.8
0.4	0.2	2.3*	2.6	2.7^	2.7^	2.6	2.7^	2.6	2.4	2.4	91.6	91.7	89.0	89.0	92.5	92.7	92.6	92.0	91.8
0.6	0.2	2.7^	2.5	2.7^	2.7^	2.5	2.6	2.4	2.4	2.3*	90.7	90.5	86.4	86.4	90.4	91.0	90.8	89.8	89.5
0.8	0.2	3.4^	2.6	2.6	2.6	2.5	2.6	[99]	2.4	2.3*	89.6	88.5	83.5	83.5	87.3	88.6	[99]	86.5	86.2
0.4	0.4	1.4*	2.6	2.8^	2.8^	2.6	2.7^	2.7^	2.5	2.4	93.6	95.8	94.2	94.3	96.3	96.6	96.6	96.0	95.9
0.6	0.4	1.7*	2.6	2.6	2.6	2.6	2.6	2.6	2.4	2.3*	92.5	94.5	92.0	92.1	94.5	95.2	95.1	94.0	93.8
0	0.5	0.6*	2.7^	2.8^	2.8^	2.7^	2.7^	[99]	2.5	2.4	96.2	98.6	98.5	98.5	99.2	99.3	[99]	99.1	99.1
0.8	0.5	1.9*	2.7^	2.7^	2.8^	2.7^	2.8^	[99]	2.5	2.4	92.6	95.3	92.4	92.5	94.4	95.5	[99]	93.9	93.7
0.6	0.6	0.8*	2.7^	2.7^	2.7^	2.7^	2.7^	2.8^	2.5	2.4	95.1	98.4	97.5	97.6	98.3	98.8	98.8	98.1	98.0
0.8	0.8	0.4*	2.7^	2.7^	2.8^	2.7^	2.8^	[99]	2.6	2.4	97.2	99.8	99.7	99.7	99.6	99.9	[99]	99.6	99.5

Table 1.84. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, $n_k \le 5$, n_k~uniform, p_C=0.3, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.5	2.5	2.6	2.6	2.6	2.6	2.7^	2.5	2.5	83.8	81.1	78.3	78.2	84.5	84.5	84.6	84.0	83.8
0.2	0	2.9^	2.6	2.6	2.6	2.6	2.5	2.5	2.5	2.4	83.1	79.7	75.9	75.8	81.9	82.0	81.8	81.4	81.2
0.4	0	3.5^	2.6	2.6	2.6	2.6	2.6	2.5	2.5	2.5	82.1	77.6	73.1	72.9	79.0	79.3	79.1	78.4	78.1
0.6	0	3.9^	2.5	2.5	2.4	2.5	2.5	2.5	2.5	2.4	81.1	76.1	70.6	70.4	75.9	76.6	76.4	75.4	75.0
0.8	0	4.5^	2.5	2.5	2.5	2.4	2.5	2.4	2.4	2.3*	80.3	73.7	67.4	67.2	72.1	73.6	73.4	71.5	71.1
0	0.2	1.6*	2.6	2.7^	2.6	2.8^	2.8^	2.9^	2.6	2.6	86.0	87.5	86.2	86.1	90.5	90.6	90.8	90.2	90.1
0.2	0.2	1.9*	2.5	2.7^	2.6	2.6	2.6	2.7^	2.5	2.5	84.8	85.7	83.3	83.2	87.6	87.9	88.0	87.3	87.0
0.4	0.2	2.4	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.4	83.9	83.8	80.3	80.2	84.9	85.4	85.2	84.4	84.2
0.6	0.2	2.8^	2.4	2.5	2.5	2.5	2.4	2.4	2.4	2.3*	83.1	82.0	77.4	77.3	81.9	82.9	82.7	81.4	81.0
0.8	0.2	3.5^	2.6	2.6	2.6	2.6	2.5	2.5	2.5	2.4	82.1	80.0	73.8	73.7	78.0	79.9	[99]	77.4	77.1
0.4	0.4	1.4*	2.6	2.7^	2.7^	2.7^	2.6	2.7^	2.6	2.5	86.6	90.7	88.7	88.7	91.6	92.3	92.4	91.3	91.1
0.6	0.4	1.8*	2.6	2.6	2.6	2.5	2.5	2.5	2.5	2.4	85.4	88.8	85.3	85.3	88.5	89.8	89.8	88.1	87.8
0	0.5	0.4*	2.6	2.7^	2.6	2.6	2.5	[99]	2.5	2.4	90.8	96.4	96.5	96.5	97.9	98.1	76.8	97.8	97.7
0.8	0.5	1.9*	2.4	2.5	2.5	2.5	2.5	2.4	2.4	2.3*	85.0	89.6	85.6	85.6	87.8	90.0	89.9	87.4	87.0
0.6	0.6	0.9*	2.7^	2.8^	2.8^	2.6	2.7^	2.8^	2.5	2.4	88.6	95.5	94.1	94.1	95.1	96.4	96.3	94.9	94.7
0.8	0.8	0.3*	2.5	2.7^	2.7^	2.5	2.6	2.8^	2.4	2.3*	91.6	99.2	98.7	98.8	98.5	99.4	[99]	98.3	98.2

Table 1.85. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, $n_k \leq 5$, n_k~uniform, p_C=0.05, δ_0=-0.02

Correlation Structure		*Type I error rate (%)*									*Power (%)*								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	3.0^	3.3^	3.4^	3.2^	3.2^	3.3^	2.6	2.5	64.3	66.5	62.7	63.4	69.7	69.7	69.8	65.6	65.1
0.2	0	2.4	2.9^	3.1^	3.2^	3.0^	3.0^	2.9^	2.4	2.3*	64.5	65.9	61.7	62.5	68.7	68.8	68.2	64.6	64.1
0.4	0	2.8^	2.9^	3.2^	3.3^	3.1^	3.1^	[99]	2.5	2.4	64.3	64.9	60.0	60.8	66.7	66.7	[99]	62.6	61.9
0.6	0	3.1^	2.9^	3.0^	3.1^	2.9^	2.9^	[99]	2.4	2.3*	63.7	63.0	57.0	57.9	63.8	64.1	[99]	59.5	58.6
0.8	0	3.5^	2.8^	2.9^	3.0^	2.7^	2.8^	[97]	2.2*	2.0*	62.9	60.1	53.2	54.2	59.3	60.2	[97]	55.1	54.1
0	0.2	1.9*	3.1^	3.3^	3.4^	3.3^	3.3^	3.5^	2.5	2.4	65.0	70.0	66.4	67.3	73.5	73.5	74.2	69.3	68.8
0.2	0.2	2.0*	2.9^	3.2^	3.4^	3.1^	3.1^	3.1^	2.4	2.3*	64.9	69.1	65.1	66.0	72.0	72.0	72.0	68.0	67.4
0.4	0.2	2.2*	2.9^	3.2^	3.3^	3.0^	3.0^	2.9^	2.4	2.2*	64.7	67.9	63.4	64.3	70.0	70.1	[99]	65.7	65.1
0.6	0.2	2.6	3.0^	3.1^	3.2^	3.0^	3.0^	[99]	2.3*	2.2*	64.5	66.4	60.6	61.6	67.3	67.7	[99]	63.0	62.2
0.8	0.2	3.0^	2.8^	2.8^	3.0^	2.7^	2.8^	[97]	2.2*	2.1*	63.0	62.8	55.9	57.0	62.0	63.0	[97]	57.5	56.6
0.4	0.4	1.6*	2.9^	3.1^	3.3^	3.0^	3.1^	[99]	2.4	2.2*	65.5	73.4	69.1	70.1	75.8	75.9	[99]	71.5	70.7
0.6	0.4	1.9*	2.9^	3.0^	3.2^	2.9^	3.0^	[99]	2.3*	2.2*	65.2	71.4	66.1	67.2	72.4	73.0	[99]	68.1	67.1
0	0.5	1.1*	3.0^	3.2^	3.4^	3.2^	3.1^	[99]	2.4	2.3*	66.7	79.1	76.0	77.0	82.3	82.5	[99]	78.8	78.0
0.8	0.5	2.0*	2.8^	2.9^	3.0^	2.8^	2.9^	[97]	2.2*	2.1*	64.3	71.3	64.7	66.0	70.4	71.9	[97]	66.0	64.9
0.6	0.6	1.0*	2.7^	3.0^	3.3^	2.8^	2.9^	[99]	2.2*	2.0*	66.5	79.9	75.5	76.6	81.0	81.7	[99]	76.8	75.8
0.8	0.8	0.5*	3.1^	3.0^	3.4^	3.0^	3.1^	[98]	2.3*	2.1*	68.2	89.7	85.9	87.3	88.6	90.9	[98]	85.2	84.1

Table 1.86. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤5, n_k~uniform, p_C=0.05, δ_0=-0.02

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.2*	3.2^	3.6^	3.8^	3.5^	3.5^	[99]	2.5	2.4	36.0	41.2	39.3	40.3	44.5	44.5	[99]	38.0	37.4
0.2	0	2.4	3.2^	3.7^	3.9^	3.5^	3.5^	[99]	2.5	2.4	35.9	40.1	37.6	38.7	42.9	42.8	[99]	36.5	35.7
0.4	0	2.5	3.1^	3.4^	3.6^	3.3^	3.3^	[99]	2.3*	2.2*	36.6	39.4	36.5	37.6	41.4	41.3	[99]	35.3	34.3
0.6	0	3.0^	3.1^	3.3^	3.4^	3.2^	3.2^	[97]	2.3*	2.1*	36.5	37.6	34.0	35.1	38.5	38.7	[96]	32.6	31.6
0.8	0	3.5^	3.0^	3.1^	3.3^	3.0^	3.0^	[86]	2.2*	2.1*	36.9	35.7	31.4	32.6	35.0	35.8	[86]	29.5	28.3
0	0.2	1.7*	3.1^	3.4^	3.6^	3.4^	3.4^	[99]	2.3*	2.2*	35.5	44.0	42.1	43.2	47.8	47.9	[99]	40.7	40.0
0.2	0.2	1.9*	3.2^	3.6^	3.8^	3.4^	3.4^	[99]	2.4	2.2*	35.6	42.9	40.6	41.8	46.0	46.0	[99]	39.4	38.5
0.4	0.2	2.1*	3.2^	3.4^	3.6^	3.3^	3.3^	[99]	2.3*	2.2*	35.5	41.3	38.5	39.7	43.5	43.6	[99]	36.9	36.0
0.6	0.2	2.4	3.0^	3.2^	3.5^	3.1^	3.2^	[97]	2.2*	2.1*	36.1	39.9	36.3	37.5	41.1	41.4	[97]	34.7	33.5
0.8	0.2	3.0^	2.8^	3.0^	3.1^	2.9^	3.0^	[88]	2.1*	1.9*	36.8	37.7	33.1	34.5	37.4	38.2	[87]	31.4	30.2
0.4	0.4	1.5*	3.1^	3.4^	3.7^	3.3^	3.3^	[99]	2.3*	2.1*	35.0	46.3	43.3	44.8	48.7	48.8	[99]	41.5	40.3
0.6	0.4	1.8*	3.2^	3.3^	3.6^	3.2^	3.2^	[97]	2.2*	2.0*	35.3	44.3	40.4	42.1	45.2	45.7	[97]	38.4	37.0
0	0.5	1.0*	3.2^	3.6^	3.8^	3.4^	3.4^	[99]	2.4	2.2*	33.8	52.0	49.5	51.1	55.5	55.7	[98]	48.3	46.9
0.8	0.5	2.0*	3.1^	3.3^	3.6^	3.2^	3.2^	[89]	2.3*	2.0*	35.4	44.4	39.1	41.1	43.5	44.9	[89]	36.7	35.0
0.6	0.6	1.0*	3.2^	3.4^	3.8^	3.4^	3.3^	[97]	2.3*	2.1*	33.7	51.9	47.7	49.7	53.1	54.0	[97]	45.4	43.7
0.8	0.8	0.4*	3.4^	3.2^	3.8^	3.4^	3.4^	[91]	2.3*	2.0*	31.6	64.4	58.2	61.7	62.5	65.9	[90]	53.9	51.2

Table 1.87. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤5, n_k~uniform, p_C=0.2, δ_0=-0.02

Correlation Structure		**Type I error rate (%)**									**Power (%)**								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.7^	3.1^	3.0^	2.9^	2.9^	3.0^	2.5	2.5	15.2	15.7	15.5	15.4	17.1	17.1	17.4	15.8	15.5
0.2	0	2.7^	2.7^	2.9^	2.9^	2.8^	2.8^	2.7^	2.5	2.4	15.9	15.2	14.6	14.5	16.0	16.1	15.9	14.8	14.5
0.4	0	3.2^	2.6	2.8^	2.8^	2.7^	2.7^	2.6	2.5	2.4	16.6	14.5	14.0	13.9	15.0	15.1	14.6	13.9	13.5
0.6	0	3.7^	2.7^	2.7^	2.7^	2.7^	2.7^	[99]	2.4	2.3*	17.4	13.9	12.9	12.9	14.0	14.2	[99]	13.0	12.6
0.8	0	4.6^	2.7^	2.7^	2.7^	2.6	2.7^	[99]	2.4	2.3*	18.7	13.3	12.4	12.3	13.0	13.4	[99]	12.1	11.7
0	0.2	1.5*	2.6	2.9^	2.9^	2.8^	2.8^	3.1^	2.4	2.3*	12.9	17.1	17.5	17.4	19.1	19.1	20.0	17.4	17.0
0.2	0.2	1.9*	2.7^	2.8^	2.8^	2.8^	2.8^	2.9^	2.5	2.4	14.2	17.0	16.8	16.8	18.3	18.3	18.5	16.8	16.4
0.4	0.2	2.3*	2.7^	2.9^	2.9^	2.7^	2.7^	[99]	2.4	2.3*	14.8	16.1	15.3	15.3	16.6	16.7	[99]	15.3	14.9
0.6	0.2	2.8^	2.6	2.7^	2.7^	2.7^	2.7^	[99]	2.4	2.3*	16.0	15.7	14.5	14.5	15.6	16.0	[99]	14.4	13.9
0.8	0.2	3.4^	2.6	2.7^	2.6	2.5	2.6	[99]	2.3*	2.2*	17.0	14.3	13.3	13.3	14.0	14.4	[99]	12.9	12.4
0.4	0.4	1.4*	2.8^	2.9^	2.9^	2.8^	2.9^	[99]	2.5	2.4	12.3	18.3	17.6	17.5	19.2	19.5	19.6	17.6	17.0
0.6	0.4	1.9*	2.5	2.8^	2.8^	2.6	2.7^	[99]	2.4	2.2*	13.7	17.7	16.6	16.6	17.7	18.1	[99]	16.2	15.7
0	0.5	0.5*	2.6	2.9^	2.9^	2.8^	2.8^	[99]	2.5	2.3*	9.3	22.8	23.3	23.3	25.4	25.5	[99]	23.7	22.9
0.8	0.5	1.8*	2.6	2.6	2.6	2.5	2.6	[99]	2.2*	2.1*	14.1	18.1	16.8	16.8	17.4	18.4	[99]	16.0	15.4
0.6	0.6	0.8*	2.7^	2.8^	2.8^	2.7^	2.7^	[99]	2.4	2.2*	10.7	21.7	20.8	20.9	21.8	22.8	[99]	20.1	19.2
0.8	0.8	0.3*	2.6	2.8^	2.8^	2.6	2.7^	[99]	2.3*	2.1*	7.9	28.8	27.4	27.8	26.8	30.0	[99]	24.6	23.2

Table 1.88. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤5, n_k~uniform, p_C=0.2, δ_0=-0.1

Correlation Structure		***Type I error rate (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.7^	2.9^	2.9^	2.9^	2.8^	2.9^	2.6	2.6
0.2	0	2.9^	2.8^	2.9^	2.8^	2.9^	2.8^	2.8^	2.6	2.6
0.4	0	3.1^	2.6	2.8^	2.7^	2.7^	2.7^	2.6	2.5	2.4
0.6	0	3.5^	2.6	2.6	2.6	2.6	2.6	[99]	2.4	2.3*
0.8	0	4.1^	2.4	2.6	2.5	2.5	2.5	[99]	2.4	2.3*
0	0.2	1.6*	2.7^	2.7^	2.7^	2.7^	2.7^	3.0^	2.5	2.5
0.2	0.2	1.9*	2.7^	2.9^	2.9^	2.7^	2.6	2.8^	2.5	2.4
0.4	0.2	2.3*	2.7^	2.8^	2.8^	2.7^	2.7^	2.7^	2.5	2.4
0.6	0.2	2.7^	2.7^	2.7^	2.7^	2.7^	2.7^	[99]	2.5	2.4
0.8	0.2	3.3^	2.6	2.7^	2.6	2.6	2.6	[99]	2.4	2.4
0.4	0.4	1.2*	2.6	2.7^	2.7^	2.7^	2.7^	2.8^	2.4	2.3*
0.6	0.4	1.7*	2.6	2.7^	2.7^	2.6	2.6	[99]	2.4	2.3*
0	0.5	0.6*	2.8^	2.7^	2.8^	2.8^	2.7^	[99]	2.6	2.4
0.8	0.5	1.8*	2.6	2.7^	2.7^	2.7^	2.7^	[99]	2.6	2.4
0.6	0.6	0.8*	2.7^	2.8^	2.9^	2.8^	2.8^	[99]	2.6	2.4
0.8	0.8	0.3*	2.9^	2.5	2.9^	2.9^	2.9^	[99]	2.7^	2.4

Power (%)

FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
99.6	99.5	99.1	99.1	99.7	99.7	99.7	99.6	99.6
99.5	99.4	98.9	98.9	99.5	99.6	99.5	99.5	99.5
99.4	99.2	98.5	98.5	99.3	99.3	[99]	99.2	99.2
99.3	99.0	98.0	98.0	99.0	99.1	[99]	98.9	98.8
99.0	98.6	97.1	97.1	98.4	98.7	[99]	98.3	98.2
99.8	99.9	99.7	99.7	99.9	99.9	99.9	99.9	99.9
99.7	99.7	99.5	99.5	99.8	99.8	99.9	99.8	99.8
99.7	99.7	99.3	99.3	99.7	99.7	99.7	99.7	99.7
99.5	99.5	99.0	99.0	99.5	99.6	[99]	99.5	99.5
99.4	99.3	98.4	98.4	99.3	99.4	[99]	99.1	99.1
99.9	99.9	99.8	99.8	99.9	100.0	100.0	99.9	99.9
99.8	99.9	99.7	99.7	99.9	99.9	[99]	99.8	99.8
100.0	100.0	100.0	100.0	100.0	100.0	[99]	100.0	100.0
99.8	99.9	99.7	99.7	99.9	99.9	[99]	99.8	99.8
99.9	100.0	100.0	100.0	100.0	100.0	[99]	100.0	100.0
100.0	100.0	100.0	100.0	100.0	100.0	[99]	100.0	100.0

Table 1.89. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤5, n_k~uniform, p_C=0.3, δ_0=-0.1

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.5	2.7^	2.6	2.6	2.6	2.7^	2.5	2.5	98.5	97.9	97.0	96.9	98.6	98.6	98.6	98.6	98.5
0.2	0	2.9^	2.6	2.7^	2.6	2.6	2.6	2.6	2.5	2.5	98.2	97.5	96.1	96.1	98.0	98.0	98.0	97.9	97.8
0.4	0	3.4^	2.5	2.6	2.5	2.6	2.6	2.5	2.5	2.4	97.9	97.0	95.2	95.1	97.3	97.4	97.3	97.1	97.1
0.6	0	3.8^	2.6	2.7^	2.6	2.5	2.5	[99]	2.4	2.3*	97.4	96.3	93.9	93.8	96.3	96.5	[99]	96.1	96.0
0.8	0	4.4^	2.4	2.6	2.5	2.5	2.5	[99]	2.4	2.3*	97.1	95.6	92.4	92.2	95.0	95.6	[99]	94.8	94.6
0	0.2	1.5*	2.5	2.5	2.5	2.5	2.5	2.7^	2.4	2.3*	99.2	99.2	98.9	98.9	99.6	99.6	99.7	99.6	99.6
0.2	0.2	1.9*	2.5	2.8^	2.7^	2.6	2.6	2.7^	2.5	2.4	98.9	98.9	98.4	98.4	99.3	99.3	99.3	99.2	99.2
0.4	0.2	2.3*	2.6	2.6	2.6	2.6	2.5	2.7^	2.5	2.4	98.6	98.6	97.7	97.7	98.8	98.9	98.9	98.7	98.7
0.6	0.2	2.8^	2.5	2.6	2.6	2.6	2.5	2.5	2.5	2.4	98.2	98.1	96.7	96.7	98.1	98.3	98.2	97.9	97.9
0.8	0.2	3.5^	2.6	2.5	2.5	2.6	2.6	[99]	2.6	2.5	97.8	97.6	95.2	95.2	97.0	97.5	[99]	96.8	96.7
0.4	0.4	1.4*	2.6	2.7^	2.7^	2.7^	2.7^	2.8^	2.6	2.5	99.3	99.6	99.3	99.3	99.7	99.7	99.7	99.7	99.6
0.6	0.4	1.8*	2.5	2.5	2.5	2.6	2.6	2.7^	2.5	2.4	99.0	99.4	98.8	98.8	99.4	99.5	99.5	99.3	99.3
0	0.5	0.4*	2.5	2.6	2.6	2.6	2.6	[99]	2.6	2.4	99.8	99.9	99.9	99.9	100.0	100.0	[99]	100.0	100.0
0.8	0.5	1.7*	2.5	2.5	2.5	2.4	2.5	[99]	2.4	2.3*	99.0	99.5	99.0	99.0	99.4	99.6	[99]	99.3	99.3
0.6	0.6	0.8*	2.5	2.6	2.6	2.6	2.6	2.8^	2.5	2.4	99.6	99.9	99.8	99.9	99.9	100.0	99.9	99.9	99.9
0.8	0.8	0.3*	2.7^	2.6	2.8^	2.7^	2.8^	[99]	2.7^	2.4	99.9	100.0	100.0	100.0	100.0	100.0	[99]	100.0	100.0

Table 1.90. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤5, n_k~uniform, p_C=0.5, δ_0=-0.1

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.4	2.4	2.3*	2.4	2.4	2.5	2.5	2.4	97.3	95.8	94.7	94.6	97.3	97.3	97.2	97.3	97.3
0.2	0	3.0^	2.3*	2.4	2.3*	2.4	2.4	2.5	2.4	2.4	96.8	95.1	93.4	93.3	96.1	96.1	96.1	96.2	96.1
0.4	0	3.6^	2.4	2.4	2.3*	2.4	2.4	2.6	2.5	2.4	96.2	94.2	91.8	91.6	94.7	94.9	94.9	94.8	94.7
0.6	0	4.3^	2.4	2.5	2.4	2.5	2.5	2.6	2.6	2.5	95.5	92.9	89.8	89.6	92.8	93.4	93.5	92.9	92.8
0.8	0	4.8^	2.4	2.5	2.4	2.5	2.5	2.6	2.6	2.5	95.0	92.0	87.6	87.4	90.9	92.0	92.2	91.0	90.7
0	0.2	1.5*	2.4	2.5	2.4	2.5	2.5	2.6	2.6	2.5	98.4	98.3	97.9	97.9	99.1	99.1	99.1	99.1	99.1
0.2	0.2	1.8*	2.4	2.5	2.4	2.4	2.4	2.5	2.5	2.4	98.1	97.8	97.1	97.0	98.5	98.6	98.6	98.5	98.5
0.4	0.2	2.5	2.4	2.4	2.4	2.5	2.5	2.6	2.5	2.4	97.4	97.1	95.7	95.6	97.4	97.7	97.7	97.5	97.4
0.6	0.2	3.0^	2.4	2.5	2.4	2.4	2.4	2.6	2.5	2.4	96.8	96.3	94.3	94.2	96.1	96.6	96.7	96.2	96.1
0.8	0.2	3.9^	2.3*	2.5	2.4	2.5	2.5	2.6	2.6	2.5	96.1	95.2	92.0	91.8	94.2	95.2	95.4	94.3	94.1
0.4	0.4	1.3*	2.4	2.5	2.5	2.4	2.5	2.6	2.6	2.4	98.6	99.0	98.7	98.6	99.2	99.3	99.3	99.3	99.2
0.6	0.4	1.9*	2.5	2.5	2.5	2.5	2.6	2.7^	2.6	2.5	98.1	98.6	97.7	97.7	98.6	98.8	98.8	98.6	98.5
0	0.5	0.4*	2.5	2.4	2.4	2.4	2.4	[99]	2.6	2.5	99.6	99.9	99.9	99.9	100.0	100.0	[99]	100.0	100.0
0.8	0.5	2.0*	2.5	2.5	2.4	2.5	2.5	2.7^	2.6	2.4	97.9	98.8	97.7	97.7	98.3	98.9	99.0	98.4	98.3
0.6	0.6	0.7*	2.4	2.3*	2.3*	2.4	2.4	2.7^	2.5	2.4	99.1	99.8	99.6	99.6	99.8	99.9	99.8	99.8	99.8
0.8	0.8	0.4*	2.7^	2.5	2.7^	2.7^	2.7^	3.3^	2.8^	2.6	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1.91. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k≤5, n_k~beta(2,3), p_C=0.1, δ_0=-0.05

Correlation Structure		**Type I error rate (%)**									**Power (%)**								
r_C	r_{CE}	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**
0	0	2.1*	3.0^	3.4^	3.4^	3.3^	3.4^	[82]	2.4	2.3*	44.4	49.3	47.1	47.9	52.1	52.1	[81]	46.1	45.2
0.2	0	2.5	3.1^	3.4^	3.5^	3.4^	3.4^	[79]	2.5	2.4	44.4	48.1	45.3	46.0	50.7	50.6	[78]	44.6	43.5
0.4	0	2.8^	3.1^	3.2^	3.3^	3.3^	3.3^	[76]	2.5	2.3*	44.2	46.6	44.0	44.8	48.6	48.6	[75]	42.6	41.4
0.6	0	2.9^	2.8^	3.0^	3.1^	3.0^	3.0^	[71]	2.2*	2.1*	44.2	45.1	41.4	42.5	45.9	46.1	[70]	40.2	38.8
0.8	0	3.3^	2.7^	2.7^	2.8^	2.7^	2.8^	[65]	2.0*	1.8*	44.6	43.2	38.5	39.5	43.1	43.7	[62]	37.5	36.0
0	0.2	1.7*	3.2^	3.6^	3.7^	3.6^	3.6^	[83]	2.5	2.3*	43.8	53.3	51.1	52.0	56.7	56.7	[81]	50.3	49.1
0.2	0.2	1.9*	3.1^	3.5^	3.6^	3.4^	3.4^	[80]	2.5	2.3*	44.2	52.3	49.5	50.5	55.0	55.0	[78]	48.8	47.6
0.4	0.2	2.1*	3.1^	3.3^	3.4^	3.3^	3.2^	[77]	2.4	2.2*	44.2	50.6	47.4	48.5	52.8	52.9	[74]	46.5	45.0
0.6	0.2	2.4	3.0^	3.0^	3.2^	3.2^	3.2^	[72]	2.3*	2.1*	43.9	48.7	44.5	45.7	49.5	49.8	[70]	43.4	41.9
0.8	0.2	2.7^	2.8^	2.8^	3.0^	2.9^	2.9^	[66]	2.2*	1.9*	44.4	46.9	41.8	43.1	46.5	47.4	[63]	40.7	39.0
0.4	0.4	1.4*	3.0^	3.2^	3.3^	3.2^	3.2^	[78]	2.3*	2.1*	43.9	57.3	53.7	55.2	59.3	59.5	[75]	52.5	50.8
0.6	0.4	1.7*	3.0^	3.1^	3.4^	3.2^	3.2^	[73]	2.4	2.1*	43.6	54.7	50.0	51.7	55.5	56.0	[71]	49.1	47.3
0	0.5	0.7*	3.2^	3.3^	3.6^	3.5^	3.4^	[83]	2.4	2.2*	42.7	64.5	62.4	64.0	68.4	68.6	[80]	61.8	60.1
0.8	0.5	1.6*	2.9^	2.8^	3.1^	2.9^	3.0^	[68]	2.2*	1.9*	43.6	55.9	49.6	51.9	55.3	56.5	[65]	48.6	46.4
0.6	0.6	0.8*	3.1^	3.1^	3.5^	3.3^	3.3^	[75]	2.4	2.1*	43.2	64.4	59.8	62.2	65.5	66.4	[71]	58.4	56.2
0.8	0.8	0.3*	3.5^	3.1^	3.8^	3.6^	3.6^	[70]	2.6	2.1*	42.3	77.8	71.4	75.2	76.5	79.2	[66]	69.6	66.4

Table 1.92. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, n_k≤5, n_k~beta(2,3), p_C=0.2, δ_0=-0.05

Correlation Structure		*Type I error rate (%)*									*Power (%)*								
r_C	r_{CE}	*FM*	*MOM*	*ICC*	*O*	*GEE Ind*	*GEE Exc*	*GEE Un*	*AFM*	*AFM CC*	*FM*	*MOM*	*ICC*	*O*	*GEE Ind*	*GEE Exc*	*GEE Un*	*AFM*	*AFM CC*
0	0	2.4	2.9^	3.2^	3.1^	3.0^	3.0^	[75]	2.6	2.5	29.3	30.6	30.3	29.9	33.1	33.1	[74]	30.3	29.6
0.2	0	2.6	2.8^	3.0^	2.9^	2.9^	3.0^	[71]	2.5	2.4	29.6	29.6	28.8	28.4	31.4	31.5	[70]	28.8	28.0
0.4	0	2.9^	2.7^	2.8^	2.7^	2.8^	2.8^	[68]	2.3*	2.2*	30.5	28.8	27.8	27.5	30.1	30.2	[67]	27.6	26.7
0.6	0	3.4^	2.7^	2.9^	2.8^	2.9^	2.9^	[65]	2.5	2.3*	31.1	27.9	26.0	25.7	28.5	28.8	[64]	26.1	25.2
0.8	0	4.0^	2.7^	2.7^	2.6	2.7^	2.8^	[65]	2.4	2.2*	31.7	26.4	24.3	24.1	26.3	26.9	[63]	24.1	23.1
0	0.2	1.7*	2.9^	3.2^	3.1^	3.1^	3.1^	[75]	2.6	2.5	28.3	34.9	34.7	34.4	38.0	37.9	[74]	35.0	34.2
0.2	0.2	1.9*	2.8^	3.1^	3.0^	3.0^	3.0^	[72]	2.5	2.4	28.3	33.3	32.7	32.5	35.4	35.5	[71]	32.6	31.7
0.4	0.2	2.2*	2.7^	2.9^	2.8^	2.9^	2.9^	[69]	2.5	2.3*	29.1	32.3	31.2	30.9	33.6	33.8	[67]	30.8	29.8
0.6	0.2	2.6	2.8^	2.9^	2.8^	2.9^	2.9^	[66]	2.5	2.3*	30.2	31.2	29.5	29.3	31.8	32.2	[65]	29.2	28.1
0.8	0.2	3.0^	2.5	2.7^	2.6	2.7^	2.7^	[65]	2.3*	2.1*	30.3	29.3	27.0	26.9	28.9	29.8	[64]	26.6	25.4
0.4	0.4	1.3*	2.7^	2.9^	2.8^	2.9^	2.9^	[70]	2.4	2.3*	27.3	37.8	36.4	36.4	39.4	39.7	[68]	36.3	34.9
0.6	0.4	1.7*	2.8^	2.9^	2.9^	2.9^	2.9^	[67]	2.5	2.3*	28.5	35.9	34.4	34.4	36.7	37.6	[65]	33.7	32.4
0	0.5	0.5*	2.9^	3.2^	3.1^	3.1^	3.1^	[75]	2.6	2.4	24.4	45.7	46.5	46.4	50.2	50.3	[73]	46.7	45.1
0.8	0.5	1.7*	2.7^	2.8^	2.8^	2.7^	2.8^	[66]	2.4	2.2*	28.2	37.0	34.1	34.4	36.4	37.7	[65]	33.4	31.8
0.6	0.6	0.8*	2.8^	2.8^	2.9^	2.9^	2.9^	[67]	2.5	2.2*	25.8	44.9	42.8	43.2	45.2	46.9	[66]	41.8	39.8
0.8	0.8	0.2*	2.9^	2.8^	3.1^	2.9^	3.0^	[66]	2.5	2.1*	22.4	58.4	54.5	56.3	55.9	60.0	[65]	51.9	49.1

Table 1.93. Empirical type I error rate, power, and/or [convergence rate (%)]: K=200, $n_k \leq 5$, $n_k \sim beta(2,3)$, $p_C=0.3$, $\delta_0=-0.05$

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.4	2.7^	2.9^	2.8^	2.7^	2.8^	[74]	2.5	2.4	24.6	24.6	24.5	24.0	26.6	26.6	[74]	25.2	24.6
0.2	0	2.8^	2.7^	2.9^	2.7^	2.8^	2.8^	[71]	2.6	2.5	25.3	23.8	23.3	22.8	25.3	25.3	[71]	23.9	23.2
0.4	0	3.2^	2.6	2.8^	2.6	2.7^	2.7^	[69]	2.5	2.3*	26.1	23.0	22.4	22.0	24.0	24.1	[68]	22.7	22.1
0.6	0	3.6^	2.6	2.6	2.5	2.7^	2.6	[67]	2.5	2.3*	26.2	21.9	20.6	20.1	22.1	22.4	[66]	21.0	20.3
0.8	0	4.1^	2.5	2.6	2.5	2.5	2.5	[67]	2.3*	2.2*	27.5	21.1	19.7	19.3	21.0	21.5	[65]	19.9	19.2
0	0.2	1.5*	2.7^	2.9^	2.8^	2.8^	2.8^	[75]	2.5	2.4	22.3	28.1	28.1	27.6	30.7	30.7	[74]	29.0	28.2
0.2	0.2	2.0*	2.7^	2.8^	2.7^	2.9^	2.9^	[72]	2.6	2.5	22.9	26.5	26.2	25.7	28.5	28.5	[71]	26.9	26.1
0.4	0.2	2.3*	2.8^	3.0^	2.8^	2.9^	2.9^	[69]	2.6	2.5	23.6	25.5	24.7	24.3	26.4	26.8	[68]	25.0	24.2
0.6	0.2	2.7^	2.5	2.7^	2.6	2.7^	2.7^	[67]	2.5	2.3*	24.8	24.7	23.5	23.0	25.2	25.7	[66]	23.8	22.9
0.8	0.2	3.1^	2.4	2.5	2.4	2.5	2.5	[67]	2.3*	2.2*	25.7	23.3	21.6	21.3	23.1	23.8	[66]	21.9	21.1
0.4	0.4	1.2*	2.5	2.7^	2.6	2.7^	2.6	[69]	2.5	2.2*	21.7	30.5	30.0	29.6	31.8	32.4	[68]	30.3	29.1
0.6	0.4	1.6*	2.7^	2.7^	2.6	2.8^	2.8^	[67]	2.5	2.3*	22.4	28.7	27.4	27.1	29.2	30.0	[66]	27.7	26.5
0	0.5	0.4*	2.6	2.8^	2.8^	2.8^	2.7^	[72]	2.5	2.3*	18.0	38.7	40.1	39.7	42.7	42.9	[71]	40.9	39.5
0.8	0.5	1.6*	2.6	2.6	2.6	2.6	2.6	[68]	2.4	2.2*	22.8	30.1	27.9	27.8	29.5	30.9	[67]	28.0	26.7
0.6	0.6	0.8*	2.7^	2.7^	2.7^	2.8^	2.8^	[67]	2.6	2.3*	19.6	36.7	35.5	35.5	37.1	38.7	[66]	35.5	33.8
0.8	0.8	0.2*	2.7^	2.7^	2.8^	2.7^	2.9^	[66]	2.5	2.2*	16.2	49.4	46.4	47.2	46.6	51.2	[66]	44.8	42.3

Table 1.94. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤5, n_k~beta(2,3), p_C=0.1, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.9^	3.1^	3.1^	3.1^	3.1^	[94]	2.5	2.5	84.5	85.2	82.3	82.7	87.2	87.2	[93]	85.1	84.8
0.2	0	2.3*	2.6	2.8^	2.8^	2.8^	2.7^	[89]	2.3*	2.2*	84.0	84.4	80.8	81.3	86.0	85.9	[88]	83.8	83.4
0.4	0	2.7^	2.7^	3.0^	3.0^	2.8^	2.8^	[83]	2.4	2.3*	83.1	83.0	79.2	79.6	84.1	84.1	[82]	81.7	81.2
0.6	0	3.0^	2.7^	2.8^	2.9^	2.7^	2.7^	[76]	2.3*	2.2*	82.5	81.6	76.9	77.5	82.1	82.3	[75]	79.6	78.9
0.8	0	3.6^	2.8^	2.9^	2.9^	2.8^	2.8^	[69]	2.4	2.2*	81.7	79.8	74.3	75.0	79.3	80.0	[67]	76.7	75.9
0	0.2	1.8*	2.9^	3.1^	3.2^	3.0^	3.0^	[94]	2.5	2.4	85.7	88.6	86.3	86.7	90.5	90.5	[93]	88.7	88.3
0.2	0.2	1.9*	2.9^	3.1^	3.1^	3.0^	3.0^	[89]	2.5	2.4	85.3	87.9	85.0	85.5	89.4	89.4	[88]	87.6	87.2
0.4	0.2	2.2*	2.9^	3.1^	3.2^	3.0^	2.9^	[84]	2.5	2.4	84.5	86.5	83.1	83.7	87.7	87.8	[82]	85.7	85.1
0.6	0.2	2.5	2.7^	2.9^	3.0^	2.9^	2.9^	[77]	2.4	2.3*	83.9	85.5	81.3	82.0	85.9	86.2	[75]	83.6	83.0
0.8	0.2	2.9^	2.7^	2.7^	2.8^	2.7^	2.8^	[71]	2.3*	2.2*	82.9	83.3	78.2	79.0	82.8	83.4	[68]	80.3	79.5
0.4	0.4	1.5*	2.9^	3.1^	3.2^	3.0^	3.0^	[84]	2.5	2.3*	86.1	90.9	88.3	88.9	91.8	92.0	[83]	90.3	89.8
0.6	0.4	1.7*	2.9^	2.9^	3.1^	2.9^	3.0^	[78]	2.4	2.2*	85.3	89.8	86.3	87.0	90.0	90.4	[76]	88.2	87.6
0	0.5	0.8*	2.8^	2.9^	3.1^	3.0^	3.0^	[93]	2.5	2.3*	89.0	95.3	94.3	94.6	96.7	96.7	[91]	95.7	95.4
0.8	0.5	1.7*	2.8^	2.7^	2.8^	2.9^	2.9^	[72]	2.4	2.2*	85.5	90.7	86.5	87.5	90.1	91.0	[70]	88.2	87.5
0.6	0.6	0.9*	3.0^	3.0^	3.2^	3.0^	3.0^	[78]	2.5	2.3*	88.2	95.2	93.3	93.8	95.3	95.8	[77]	94.2	93.8
0.8	0.8	0.3*	3.2^	2.8^	3.3^	3.2^	3.2^	[74]	2.6	2.3*	91.3	98.8	98.0	98.3	98.6	99.0	[71]	98.0	97.7

Table 1.95. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤5, n_k~beta(2,3), p_C=0.2, δ_0=-0.05

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.4	2.6	2.9^	2.9^	2.8^	2.8^	[96]	2.6	2.5	62.7	62.1	60.4	60.2	65.1	65.0	[96]	63.5	63.0
0.2	0	2.6	2.7^	2.8^	2.7^	2.7^	2.6	[93]	2.4	2.4	62.4	60.8	58.2	58.0	62.8	63.0	[92]	61.1	60.5
0.4	0	3.0^	2.5	2.8^	2.7^	2.6	2.6	[89]	2.4	2.4	62.4	59.3	56.2	56.1	60.7	60.8	[89]	59.1	58.5
0.6	0	3.5^	2.6	2.8^	2.7^	2.7^	2.7^	[85]	2.5	2.4	61.6	57.3	53.3	53.1	57.4	57.9	[84]	55.7	55.0
0.8	0	3.9^	2.5	2.6	2.5	2.5	2.5	[80]	2.3*	2.2*	60.9	54.7	50.2	50.0	53.7	54.7	[78]	52.1	51.3
0	0.2	1.6*	2.7^	2.9^	2.9^	2.8^	2.8^	[96]	2.6	2.5	64.4	69.1	67.6	67.5	72.5	72.5	[96]	70.9	70.4
0.2	0.2	2.0*	2.7^	2.9^	2.8^	2.8^	2.8^	[93]	2.5	2.4	63.2	66.8	64.4	64.4	69.2	69.3	[93]	67.5	66.9
0.4	0.2	2.1*	2.5	2.7^	2.7^	2.6	2.6	[90]	2.4	2.2*	62.9	64.9	61.9	61.9	66.2	66.6	[89]	64.6	64.0
0.6	0.2	2.5	2.5	2.6	2.6	2.6	2.6	[85]	2.3*	2.3*	62.5	62.8	58.9	58.9	63.1	63.8	[84]	61.4	60.6
0.8	0.2	3.0^	2.5	2.6	2.5	2.5	2.5	[80]	2.4	2.2*	62.6	60.9	55.9	56.0	59.8	61.1	[79]	58.2	57.4
0.4	0.4	1.3*	2.6	2.6	2.6	2.7^	2.7^	[90]	2.4	2.3*	64.3	73.4	70.7	70.8	74.7	75.4	[89]	73.1	72.3
0.6	0.4	1.6*	2.5	2.6	2.6	2.7^	2.6	[85]	2.4	2.3*	64.3	71.1	67.6	67.7	71.4	72.3	[85]	69.7	68.9
0	0.5	0.5*	2.7^	2.8^	2.8^	2.9^	2.9^	[92]	2.6	2.4	66.5	82.7	82.8	82.9	86.1	86.4	[92]	85.0	84.4
0.8	0.5	1.5*	2.5	2.6	2.6	2.5	2.6	[81]	2.3*	2.1*	64.0	72.6	68.1	68.4	71.3	73.3	[80]	69.5	68.6
0.6	0.6	0.8*	2.8^	2.9^	2.9^	2.9^	2.9^	[86]	2.6	2.4	66.2	82.0	79.3	79.7	81.9	83.4	[85]	80.4	79.6
0.8	0.8	0.3*	2.8^	2.6	2.8^	2.8^	2.9^	[82]	2.6	2.4	67.9	92.6	90.5	91.1	91.0	93.4	[80]	89.9	89.1

Table 1.96. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤5, n_k~beta(2,3), p_C=0.3, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.6	2.8^	2.7^	2.7^	2.7^	[97]	2.6	2.5	52.7	51.4	49.8	49.4	54.1	54.1	[97]	53.2	52.8
0.2	0	2.8^	2.5	2.7^	2.6	2.6	2.6	[95]	2.4	2.3*	52.8	49.9	47.5	47.2	51.7	51.8	[95]	50.7	50.2
0.4	0	3.2^	2.6	2.7^	2.7^	2.6	2.6	[93]	2.5	2.4	52.2	48.0	44.9	44.6	48.8	49.1	[92]	47.9	47.3
0.6	0	3.7^	2.5	2.7^	2.6	2.6	2.5	[90]	2.5	2.4	52.8	46.8	43.5	43.2	46.8	47.4	[89]	45.9	45.3
0.8	0	4.3^	2.6	2.6	2.5	2.6	2.6	[86]	2.5	2.4	52.4	44.5	40.6	40.3	43.7	44.5	[85]	42.8	42.2
0	0.2	1.5*	2.5	2.7^	2.7^	2.6	2.6	[97]	2.5	2.4	52.8	58.1	57.5	57.2	61.9	61.9	[97]	60.9	60.3
0.2	0.2	1.8*	2.6	2.6	2.6	2.5	2.6	[95]	2.4	2.3*	52.6	56.0	54.6	54.3	58.6	58.7	[95]	57.6	57.0
0.4	0.2	2.1*	2.5	2.6	2.5	2.5	2.5	[93]	2.4	2.3*	52.9	54.7	51.8	51.5	55.7	56.2	[93]	54.7	54.1
0.6	0.2	2.7^	2.6	2.6	2.5	2.6	2.6	[90]	2.5	2.4	52.6	52.4	48.8	48.6	52.3	53.1	[90]	51.3	50.6
0.8	0.2	3.2^	2.5	2.6	2.6	2.5	2.5	[86]	2.4	2.3*	52.6	49.9	45.5	45.2	48.8	50.0	[85]	47.8	47.1
0.4	0.4	1.3*	2.6	2.7^	2.7^	2.8^	2.7^	[93]	2.6	2.5	53.1	63.1	60.9	60.8	64.6	65.3	[93]	63.6	62.8
0.6	0.4	1.7*	2.6	2.6	2.6	2.5	2.6	[90]	2.4	2.3*	53.1	60.7	57.2	57.0	60.4	61.9	[90]	59.4	58.6
0	0.5	0.4*	2.7^	2.7^	2.7^	2.7^	2.7^	[92]	2.6	2.4	54.0	74.7	75.7	75.7	79.3	79.4	[91]	78.6	77.8
0.8	0.5	1.7*	2.5	2.6	2.6	2.6	2.5	[87]	2.5	2.3*	52.8	62.1	57.7	57.7	60.2	62.6	[86]	59.2	58.2
0.6	0.6	0.7*	2.6	2.6	2.6	2.6	2.6	[90]	2.5	2.3*	53.4	72.2	70.1	70.1	72.1	74.1	[90]	71.1	70.1
0.8	0.8	0.2*	2.7^	2.6	2.7^	2.7^	2.7^	[87]	2.6	2.4	53.6	86.0	83.8	84.2	83.2	87.3	[87]	82.4	81.4

Table 1.97. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k≤5, n_k~beta(2,3), p_C=0.1, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.7^	3.0^	3.0^	2.8^	2.9^	[99]	2.5	2.4	98.9	98.8	98.0	98.1	99.1	99.1	[99]	99.0	98.9
0.2	0	2.5	2.7^	2.8^	2.9^	2.8^	2.8^	[97]	2.4	2.4	98.8	98.6	97.8	97.9	98.9	98.9	[97]	98.8	98.7
0.4	0	2.8^	2.6	2.7^	2.7^	2.7^	2.7^	[94]	2.4	2.4	98.6	98.4	97.4	97.5	98.6	98.6	[94]	98.4	98.3
0.6	0	3.2^	2.7^	2.7^	2.8^	2.7^	2.7^	[89]	2.4	2.3*	98.4	98.2	96.7	96.8	98.2	98.2	[88]	97.9	97.9
0.8	0	3.5^	2.6	2.7^	2.7^	2.5	2.6	[81]	2.3*	2.2*	98.1	97.8	96.0	96.1	97.6	97.7	[79]	97.1	97.1
0	0.2	1.8*	2.7^	2.9^	3.0^	2.9^	2.8^	[99]	2.5	2.4	99.2	99.4	99.0	99.0	99.6	99.6	[99]	99.5	99.4
0.2	0.2	2.0*	2.7^	3.0^	3.1^	2.9^	2.9^	[97]	2.6	2.5	99.1	99.2	98.8	98.8	99.4	99.4	[97]	99.3	99.3
0.4	0.2	2.2*	2.8^	2.8^	2.9^	2.8^	2.8^	[94]	2.5	2.4	99.0	99.1	98.5	98.5	99.2	99.2	[94]	99.1	99.1
0.6	0.2	2.4	2.6	2.6	2.7^	2.6	2.6	[90]	2.3*	2.3*	98.8	98.9	98.0	98.1	98.9	99.0	[89]	98.7	98.7
0.8	0.2	3.0^	2.7^	2.8^	2.9^	2.7^	2.7^	[82]	2.4	2.4	98.4	98.5	97.2	97.3	98.3	98.5	[80]	98.1	98.0
0.4	0.4	1.5*	2.7^	2.8^	2.9^	2.8^	2.8^	[95]	2.4	2.3*	99.4	99.6	99.4	99.4	99.7	99.7	[94]	99.6	99.6
0.6	0.4	1.7*	2.7^	2.7^	2.9^	2.7^	2.8^	[90]	2.3*	2.2*	99.2	99.6	99.1	99.2	99.6	99.6	[89]	99.5	99.5
0	0.5	0.8*	2.8^	2.9^	3.0^	3.0^	3.0^	[97]	2.6	2.5	99.8	99.9	99.9	99.9	100.0	100.0	[96]	100.0	100.0
0.8	0.5	1.8*	2.8^	2.8^	2.9^	2.8^	2.9^	[84]	2.5	2.4	99.2	99.6	99.1	99.2	99.5	99.6	[82]	99.4	99.4
0.6	0.6	0.9*	2.8^	2.8^	2.9^	2.8^	2.8^	[91]	2.4	2.3*	99.6	99.9	99.8	99.8	99.9	99.9	[90]	99.9	99.9
0.8	0.8	0.4*	3.0^	2.8^	3.1^	3.0^	3.1^	[85]	2.6	2.4	99.9	100.0	100.0	100.0	100.0	100.0	[84]	100.0	100.0

Table 1.98. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, $n_k \leq 5$, n_k~beta(2,3), p_C=0.2, δ_0=-0.05

Correlation Structure		**Type I error rate (%)**									**Power (%)**								
r_C	r_{CE}	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**	**FM**	**MOM**	**ICC**	**O**	**GEE Ind**	**GEE Exc**	**GEE Un**	**AFM**	**AFM CC**
0	0	2.4	2.6	2.8^	2.8^	2.7^	2.7^	[99]	2.5	2.5	90.4	89.4	87.6	87.6	91.1	91.1	[99]	90.6	90.4
0.2	0	2.7^	2.6	2.6	2.6	2.6	2.6	[99]	2.4	2.4	89.7	88.3	85.8	85.8	89.6	89.7	[99]	89.0	88.8
0.4	0	3.2^	2.6	2.8^	2.8^	2.7^	2.7^	[98]	2.5	2.5	89.3	87.4	84.4	84.4	88.1	88.3	[98]	87.5	87.3
0.6	0	3.3^	2.4	2.6	2.5	2.5	2.5	[96]	2.3*	2.3*	88.3	85.9	82.2	82.2	85.9	86.3	[95]	85.2	84.9
0.8	0	3.9^	2.5	2.6	2.5	2.5	2.5	[91]	2.4	2.3*	87.2	83.9	79.1	79.1	82.7	83.6	[91]	82.0	81.7
0	0.2	1.7*	2.7^	2.8^	2.8^	2.7^	2.8^	[99]	2.5	2.5	92.2	93.4	92.3	92.3	94.9	94.9	[99]	94.5	94.4
0.2	0.2	1.9*	2.5	2.6	2.6	2.6	2.6	[99]	2.4	2.4	91.1	92.0	90.5	90.5	93.2	93.3	[99]	92.7	92.6
0.4	0.2	2.2*	2.5	2.6	2.6	2.6	2.6	[98]	2.5	2.4	90.9	91.5	89.3	89.3	92.1	92.3	[98]	91.6	91.4
0.6	0.2	2.6	2.5	2.6	2.6	2.5	2.6	[96]	2.4	2.3*	89.7	89.9	86.9	86.9	89.8	90.3	[96]	89.2	89.0
0.8	0.2	3.3^	2.6	2.6	2.6	2.6	2.6	[92]	2.5	2.4	88.9	88.0	84.1	84.2	87.3	88.1	[91]	86.6	86.3
0.4	0.4	1.3*	2.6	2.7^	2.7^	2.6	2.6	[98]	2.5	2.4	92.9	95.5	94.3	94.3	96.0	96.2	[98]	95.7	95.5
0.6	0.4	1.7*	2.6	2.6	2.6	2.6	2.6	[96]	2.4	2.3*	91.9	94.3	92.4	92.5	94.2	94.8	[96]	93.9	93.7
0	0.5	0.5*	2.5	2.6	2.6	2.7^	2.6	[98]	2.5	2.4	95.5	98.5	98.4	98.4	99.1	99.1	[98]	99.0	98.9
0.8	0.5	1.7*	2.4	2.5	2.5	2.5	2.5	[92]	2.4	2.3*	91.9	95.0	92.7	92.8	94.3	95.2	[92]	93.9	93.6
0.6	0.6	0.8*	2.6	2.7^	2.7^	2.7^	2.6	[96]	2.5	2.4	94.5	98.1	97.4	97.5	98.1	98.4	[96]	97.9	97.8
0.8	0.8	0.2*	2.6	2.6	2.8^	2.7^	2.6	[93]	2.6	2.3*	96.7	99.8	99.7	99.7	99.7	99.8	[93]	99.6	99.6

Table 1.99. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k≤5, n_k~beta(2,3), p_C=0.3, δ_0=-0.05

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.6	2.6	2.7^	2.7^	2.7^	2.7^	[99]	2.6	2.6	82.2	80.3	78.2	78.1	82.8	82.8	[99]	82.4	82.2
0.2	0	2.9^	2.6	2.6	2.6	2.6	2.6	[99]	2.6	2.5	81.6	78.8	75.9	75.8	80.6	80.7	[99]	80.1	79.9
0.4	0	3.0^	2.4	2.4	2.4	2.4	2.4	[99]	2.3*	2.3*	80.5	76.9	73.5	73.3	77.7	78.0	[99]	77.3	77.0
0.6	0	3.5^	2.4	2.5	2.4	2.4	2.4	[98]	2.3*	2.3*	79.8	75.2	70.9	70.7	75.0	75.7	[98]	74.6	74.2
0.8	0	4.4^	2.6	2.6	2.6	2.6	2.6	[96]	2.6	2.5	79.3	73.3	68.2	68.0	72.0	73.2	[96]	71.5	71.1
0	0.2	1.6*	2.6	2.7^	2.6	2.7^	2.6	[99]	2.6	2.5	84.3	86.4	85.2	85.1	88.9	88.9	[99]	88.6	88.4
0.2	0.2	1.9*	2.5	2.6	2.6	2.7^	2.7^	[99]	2.6	2.5	83.4	85.0	83.0	82.9	86.6	86.7	[99]	86.3	86.0
0.4	0.2	2.3*	2.6	2.6	2.5	2.6	2.6	[99]	2.5	2.5	82.5	83.2	80.7	80.6	84.1	84.6	[99]	83.7	83.4
0.6	0.2	2.6	2.5	2.6	2.5	2.5	2.5	[98]	2.4	2.4	81.7	81.4	77.7	77.6	81.2	82.0	[98]	80.8	80.5
0.8	0.2	3.3^	2.5	2.5	2.5	2.5	2.5	[96]	2.5	2.4	80.8	79.1	74.4	74.4	77.9	79.1	[96]	77.4	77.0
0.4	0.4	1.2*	2.5	2.5	2.5	2.5	2.6	[99]	2.4	2.4	85.2	90.1	88.3	88.3	90.8	91.3	[99]	90.6	90.3
0.6	0.4	1.6*	2.5	2.5	2.5	2.5	2.5	[98]	2.4	2.3*	84.3	88.3	85.8	85.8	88.2	89.1	[98]	87.9	87.6
0	0.5	0.5*	2.6	2.7^	2.7^	2.7^	2.6	[98]	2.6	2.5	88.9	95.8	96.1	96.1	97.4	97.5	[98]	97.3	97.2
0.8	0.5	1.7*	2.5	2.6	2.5	2.6	2.6	[96]	2.5	2.4	84.1	89.4	86.2	86.2	88.1	89.7	[96]	87.7	87.3
0.6	0.6	0.8*	2.6	2.6	2.6	2.6	2.6	[98]	2.5	2.4	87.4	95.2	93.9	94.0	94.9	95.8	[98]	94.7	94.5
0.8	0.8	0.2*	2.6	2.4	2.5	2.5	2.5	[97]	2.4	2.3*	90.4	99.1	98.7	98.7	98.5	99.2	[96]	98.4	98.3

Table 1.100. Empirical type I error rate, power, and/or [convergence rate (%)]: K=1000, n_k≤5, n_k~beta(2,3), p_C=0.05, δ_0=-0.02

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.3*	3.0^	3.2^	3.3^	3.1^	3.1^	[97]	2.5	2.5	63.3	65.8	62.6	63.5	68.3	68.4	[97]	64.7	64.1
0.2	0	2.4	3.0^	3.2^	3.4^	3.0^	3.0^	[94]	2.4	2.3*	63.6	65.3	61.6	62.5	67.7	67.7	[94]	63.9	63.2
0.4	0	2.7^	2.9^	3.0^	3.2^	3.0^	3.0^	[88]	2.4	2.3*	62.5	63.4	59.7	60.5	65.1	65.1	[88]	61.2	60.5
0.6	0	2.9^	2.9^	2.9^	3.1^	2.8^	2.9^	[80]	2.3*	2.2*	62.4	62.1	57.4	58.3	62.9	63.0	[80]	59.0	58.2
0.8	0	3.3^	2.6	2.7^	2.8^	2.6	2.7^	[70]	2.1*	2.0*	61.5	59.6	53.9	54.9	58.8	59.5	[69]	55.0	54.1
0	0.2	1.9*	3.0^	3.2^	3.3^	3.1^	3.1^	[97]	2.5	2.4	63.7	68.9	66.3	67.1	71.7	71.7	[97]	68.0	67.3
0.2	0.2	2.0*	3.0^	3.2^	3.3^	3.1^	3.1^	[94]	2.5	2.4	63.5	68.1	64.8	65.7	70.4	70.4	[93]	66.5	65.9
0.4	0.2	2.3*	2.9^	3.1^	3.3^	3.1^	3.1^	[88]	2.5	2.4	63.3	67.0	63.3	64.1	68.6	68.7	[88]	64.9	64.2
0.6	0.2	2.4	2.7^	2.9^	3.0^	2.8^	2.8^	[80]	2.3*	2.2*	62.8	65.1	60.6	61.6	65.8	66.1	[79]	61.9	61.1
0.8	0.2	2.9^	2.7^	2.8^	3.0^	2.7^	2.8^	[70]	2.2*	2.1*	62.3	62.7	57.2	58.4	62.2	62.9	[69]	58.3	57.4
0.4	0.4	1.5*	2.9^	2.9^	3.1^	3.0^	2.9^	[88]	2.3*	2.2*	64.1	72.4	68.8	69.8	74.1	74.3	[87]	70.3	69.5
0.6	0.4	1.8*	2.8^	2.9^	3.1^	3.0^	2.9^	[80]	2.3*	2.2*	63.3	70.0	65.8	66.9	70.7	71.1	[79]	66.9	65.9
0	0.5	1.0*	2.9^	3.0^	3.2^	3.2^	3.1^	[96]	2.4	2.3*	65.2	78.5	76.3	77.2	81.4	81.4	[95]	77.9	77.1
0.8	0.5	1.8*	2.7^	2.7^	2.9^	2.7^	2.7^	[72]	2.2*	2.1*	63.1	71.0	65.2	66.5	70.1	71.5	[70]	66.1	64.9
0.6	0.6	1.1*	2.9^	3.0^	3.3^	2.9^	3.0^	[81]	2.3*	2.2*	64.8	78.7	74.9	76.1	79.4	80.1	[80]	75.8	74.7
0.8	0.8	0.5*	3.2^	3.0^	3.4^	3.1^	3.2^	[74]	2.5	2.2*	66.9	89.2	85.7	87.1	88.2	90.0	[73]	85.1	84.0

Table 1.101. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, $n_k \leq 5$, n_k~beta(2,3), p_C=0.05, δ_0=-0.02

Correlation Structure		Type I error rate (%)									Power (%)								
r_C	r_{CE}	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC	FM	MOM	ICC	O	GEE Ind	GEE Exc	GEE Un	AFM	AFM CC
0	0	2.3*	3.3^	3.5^	3.7^	3.4^	3.4^	[95]	2.6	2.4	35.0	40.2	38.4	39.5	42.8	42.8	[94]	37.0	36.1
0.2	0	2.3*	3.0^	3.4^	3.6^	3.2^	3.2^	[90]	2.4	2.3*	35.2	39.4	37.5	38.6	41.6	41.6	[90]	35.9	35.1
0.4	0	2.6	3.1^	3.3^	3.5^	3.2^	3.2^	[85]	2.4	2.3*	35.2	38.5	36.3	37.4	39.9	39.9	[85]	34.4	33.4
0.6	0	2.8^	2.9^	3.2^	3.4^	3.0^	3.0^	[79]	2.3*	2.1*	35.3	37.0	34.1	35.4	37.5	37.8	[78]	32.3	31.4
0.8	0	3.2^	2.9^	2.9^	3.0^	2.8^	2.9^	[69]	2.2*	2.0*	36.1	35.2	31.6	32.8	35.0	35.4	[69]	30.0	28.8
0	0.2	1.8*	3.2^	3.6^	3.8^	3.6^	3.6^	[95]	2.6	2.4	34.5	42.6	41.0	42.2	45.6	45.6	[94]	39.5	38.6
0.2	0.2	2.0*	3.2^	3.4^	3.6^	3.5^	3.5^	[91]	2.5	2.4	34.5	42.2	39.8	41.1	44.5	44.5	[90]	38.4	37.3
0.4	0.2	2.2*	3.2^	3.4^	3.6^	3.4^	3.4^	[85]	2.5	2.3*	35.0	41.0	38.4	39.6	42.6	42.5	[85]	36.9	35.7
0.6	0.2	2.5	3.0^	3.1^	3.3^	3.2^	3.2^	[78]	2.3*	2.2*	35.0	39.2	36.1	37.4	40.0	40.3	[78]	34.4	33.2
0.8	0.2	2.7^	2.8^	2.7^	2.9^	2.7^	2.8^	[70]	2.1*	1.9*	35.6	37.2	33.6	35.0	36.9	37.4	[69]	31.5	30.2
0.4	0.4	1.5*	3.0^	3.3^	3.5^	3.2^	3.2^	[86]	2.3*	2.1*	34.4	45.5	42.9	44.5	47.4	47.5	[85]	41.2	39.9
0.6	0.4	1.7*	3.0^	3.2^	3.4^	3.1^	3.1^	[80]	2.2*	2.0*	34.5	43.5	40.2	41.8	44.4	44.8	[79]	38.2	36.9
0	0.5	1.0*	3.3^	3.4^	3.7^	3.5^	3.5^	[93]	2.5	2.3*	32.6	50.9	49.3	50.9	54.2	54.2	[91]	47.5	46.1
0.8	0.5	1.7*	3.1^	3.0^	3.3^	3.0^	3.1^	[72]	2.2*	2.0*	34.0	43.7	38.9	40.9	43.2	44.1	[70]	36.9	35.1
0.6	0.6	1.0*	3.3^	3.2^	3.5^	3.4^	3.3^	[81]	2.4	2.1*	32.6	51.2	47.6	49.8	52.2	52.9	[80]	45.0	43.1
0.8	0.8	0.4*	3.3^	3.0^	3.6^	3.2^	3.3^	[75]	2.3*	2.0*	30.3	63.4	58.1	61.7	62.2	64.8	[73]	54.3	51.6

Table 1.102. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤5, n_k~beta(2,3), p_C=0.2, δ_0=-0.02

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.3*	2.6	2.8^	2.8^	2.7^	2.7^	[96]	2.4	2.3*	14.7	15.2	15.2	15.1	16.3	16.3	[96]	15.2	14.9
0.2	0	2.7^	2.6	2.8^	2.7^	2.8^	2.8^	[93]	2.5	2.4	15.2	14.8	14.4	14.3	15.5	15.5	[93]	14.4	14.1
0.4	0	3.2^	2.7^	2.9^	2.8^	2.8^	2.9^	[89]	2.6	2.5	16.2	14.4	14.2	14.1	15.0	15.0	[89]	13.9	13.6
0.6	0	3.5^	2.6	2.7^	2.7^	2.6	2.7^	[84]	2.4	2.3*	16.9	14.0	13.3	13.2	14.0	14.2	[84]	13.0	12.7
0.8	0	4.0^	2.6	2.6	2.6	2.5	2.6	[79]	2.3*	2.2*	17.8	13.3	12.4	12.3	13.1	13.3	[78]	12.2	11.8
0	0.2	1.6*	2.7^	2.9^	2.8^	2.9^	2.9^	[96]	2.6	2.4	12.8	16.9	17.1	17.0	18.4	18.4	[96]	17.1	16.7
0.2	0.2	1.9*	2.7^	2.9^	2.9^	2.8^	2.8^	[93]	2.5	2.4	13.6	16.5	16.4	16.3	17.3	17.4	[93]	16.1	15.7
0.4	0.2	2.2*	2.7^	2.8^	2.8^	2.7^	2.7^	[89]	2.4	2.3*	14.1	15.7	15.4	15.4	16.2	16.3	[89]	15.1	14.6
0.6	0.2	2.6	2.6	2.6	2.6	2.6	2.7^	[85]	2.4	2.3*	15.5	15.5	14.9	14.8	15.5	15.9	[84]	14.5	14.1
0.8	0.2	3.0^	2.5	2.6	2.6	2.5	2.5	[79]	2.4	2.2*	16.2	14.5	13.6	13.6	14.2	14.6	[79]	13.3	12.8
0.4	0.4	1.3*	2.6	2.7^	2.7^	2.7^	2.7^	[89]	2.4	2.3*	12.2	18.3	18.2	18.1	19.1	19.2	[89]	17.8	17.2
0.6	0.4	1.7*	2.6	2.7^	2.7^	2.7^	2.6	[85]	2.4	2.3*	13.2	17.2	16.7	16.7	17.6	17.9	[85]	16.4	15.8
0	0.5	0.5*	2.7^	2.9^	2.9^	2.9^	2.8^	[92]	2.6	2.4	8.8	22.1	22.9	22.9	24.4	24.4	[91]	22.6	21.9
0.8	0.5	1.7*	2.6	2.8^	2.8^	2.7^	2.7^	[81]	2.4	2.3*	13.1	17.5	16.4	16.4	17.2	17.8	[80]	16.1	15.4
0.6	0.6	0.8*	2.6	2.8^	2.8^	2.8^	2.7^	[85]	2.5	2.3*	9.8	20.8	20.3	20.4	20.8	21.7	[85]	19.4	18.5
0.8	0.8	0.3*	2.7^	2.8^	2.9^	2.7^	2.8^	[81]	2.5	2.2*	7.0	28.3	26.8	27.2	26.6	29.0	[81]	24.8	23.4

Table 1.103. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, $n_k \leq 5$, n_k~beta(2,3), p_C=0.2, δ_0=-0.1

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.4	2.6	2.7^	2.6	2.7^	2.7^	[96]	2.5	2.5	99.5	99.4	99.1	99.1	99.6	99.6	[96]	99.6	99.5
0.2	0	2.8^	2.7^	2.8^	2.7^	2.8^	2.8^	[93]	2.6	2.5	99.4	99.3	98.8	98.9	99.4	99.4	[93]	99.4	99.4
0.4	0	2.9^	2.5	2.6	2.6	2.6	2.6	[90]	2.4	2.3*	99.2	99.1	98.5	98.5	99.2	99.2	[89]	99.1	99.1
0.6	0	3.3^	2.4	2.5	2.5	2.5	2.4	[85]	2.3*	2.2*	99.2	99.0	98.1	98.1	99.0	99.0	[84]	98.9	98.8
0.8	0	3.8^	2.5	2.5	2.5	2.4	2.5	[81]	2.3*	2.2*	99.1	98.7	97.4	97.4	98.5	98.7	[78]	98.4	98.3
0	0.2	1.6*	2.7^	2.8^	2.8^	2.7^	2.7^	[96]	2.6	2.5	99.7	99.8	99.7	99.7	99.9	99.9	[96]	99.9	99.9
0.2	0.2	1.9*	2.7^	2.7^	2.7^	2.8^	2.7^	[94]	2.5	2.5	99.6	99.7	99.5	99.5	99.8	99.8	[93]	99.8	99.8
0.4	0.2	2.2*	2.6	2.7^	2.7^	2.6	2.6	[90]	2.5	2.4	99.6	99.6	99.3	99.3	99.7	99.7	[89]	99.6	99.6
0.6	0.2	2.5	2.5	2.6	2.6	2.6	2.6	[86]	2.4	2.3*	99.5	99.5	99.0	99.0	99.5	99.6	[84]	99.5	99.4
0.8	0.2	3.0^	2.5	2.6	2.6	2.5	2.6	[81]	2.4	2.3*	99.3	99.2	98.5	98.5	99.2	99.3	[79]	99.1	99.0
0.4	0.4	1.4*	2.7^	2.8^	2.8^	2.8^	2.8^	[90]	2.6	2.5	99.8	99.9	99.9	99.9	99.9	100.0	[89]	99.9	99.9
0.6	0.4	1.7*	2.8^	2.7^	2.7^	2.7^	2.7^	[86]	2.5	2.4	99.7	99.8	99.7	99.7	99.8	99.9	[85]	99.8	99.8
0	0.5	0.6*	2.7^	2.8^	2.9^	2.9^	2.8^	[93]	2.6	2.5	100.0	100.0	100.0	100.0	100.0	100.0	[91]	100.0	100.0
0.8	0.5	1.6*	2.6	2.7^	2.7^	2.6	2.7^	[82]	2.5	2.4	99.7	99.9	99.7	99.8	99.9	99.9	[80]	99.9	99.8
0.6	0.6	0.8*	2.7^	2.7^	2.8^	2.8^	2.8^	[86]	2.6	2.4	99.9	100.0	100.0	100.0	100.0	100.0	[85]	100.0	100.0
0.8	0.8	0.3*	2.9^	2.6	3.0^	2.8^	3.0^	[82]	2.7^	2.5	100.0	100.0	100.0	100.0	100.0	100.0	[81]	100.0	100.0

Table 1.104. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, n_k≤5, n_k~beta(2,3), p_C=0.3, δ_0=-0.1

Correlation Structure		***Type I error rate (%)***									***Power (%)***								
r_C	***r_{CE}***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.5	2.6	2.7^	2.6	2.6	2.6	[97]	2.6	2.5	98.2	97.7	97.0	96.9	98.3	98.3	[97]	98.2	98.1
0.2	0	2.8^	2.5	2.6	2.5	2.5	2.5	[95]	2.5	2.4	97.9	97.3	96.2	96.2	97.8	97.8	[95]	97.6	97.6
0.4	0	3.2^	2.5	2.8^	2.7^	2.6	2.6	[93]	2.5	2.4	97.5	96.7	95.3	95.2	96.9	97.0	[93]	96.8	96.7
0.6	0	3.6^	2.4	2.6	2.5	2.5	2.5	[90]	2.5	2.4	97.3	96.3	94.5	94.4	96.3	96.5	[89]	96.1	96.0
0.8	0	4.2^	2.5	2.7^	2.6	2.6	2.6	[87]	2.5	2.4	96.7	95.3	92.9	92.8	94.8	95.3	[85]	94.6	94.5
0	0.2	1.6*	2.5	2.7^	2.6	2.6	2.5	[97]	2.5	2.4	98.9	99.0	98.8	98.8	99.4	99.4	[97]	99.4	99.4
0.2	0.2	1.9*	2.5	2.6	2.6	2.6	2.5	[95]	2.5	2.4	98.6	98.7	98.3	98.3	99.1	99.1	[95]	99.0	99.0
0.4	0.2	2.2*	2.5	2.6	2.6	2.5	2.5	[93]	2.4	2.4	98.4	98.4	97.6	97.6	98.6	98.7	[92]	98.5	98.5
0.6	0.2	2.6	2.5	2.6	2.6	2.5	2.5	[90]	2.4	2.3*	98.0	98.0	96.9	96.8	98.0	98.2	[90]	97.9	97.8
0.8	0.2	3.2^	2.5	2.5	2.5	2.6	2.6	[87]	2.5	2.4	97.6	97.5	95.9	95.9	97.1	97.5	[85]	97.0	96.8
0.4	0.4	1.3*	2.5	2.6	2.6	2.6	2.6	[93]	2.5	2.4	99.1	99.5	99.3	99.3	99.6	99.7	[93]	99.6	99.6
0.6	0.4	1.7*	2.6	2.7^	2.6	2.6	2.7^	[91]	2.6	2.5	98.9	99.3	98.8	98.8	99.3	99.4	[90]	99.3	99.3
0	0.5	0.5*	2.7^	2.7^	2.7^	2.7^	2.7^	[92]	2.6	2.5	99.7	99.9	99.9	99.9	100.0	100.0	[91]	100.0	100.0
0.8	0.5	1.7*	2.6	2.6	2.6	2.6	2.7^	[87]	2.5	2.4	98.8	99.4	98.9	99.0	99.3	99.5	[86]	99.3	99.2
0.6	0.6	0.7*	2.6	2.5	2.6	2.6	2.7^	[91]	2.6	2.4	99.5	99.9	99.8	99.8	99.9	99.9	[90]	99.9	99.9
0.8	0.8	0.3*	2.7^	2.6	2.8^	2.8^	2.8^	[87]	2.8^	2.5	99.9	100.0	100.0	100.0	100.0	100.0	[87]	100.0	100.0

Table 1.105. Empirical type I error rate, power, and/or [convergence rate (%)]: K=500, $n_k \leq 5$, n_k~beta(2,3), p_C=0.5, δ_0=-0.1

Correlation Structure		**Type I error rate (%)**									**Power (%)**								
r_C	r_{CE}	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***	***FM***	***MOM***	***ICC***	***O***	***GEE Ind***	***GEE Exc***	***GEE Un***	***AFM***	***AFM CC***
0	0	2.6	2.5	2.5	2.4	2.4	2.4	[97]	2.5	2.5	96.7	95.6	94.5	94.4	96.7	96.7	[97]	96.7	96.7
0.2	0	2.9^	2.4	2.5	2.4	2.4	2.4	[96]	2.5	2.4	96.2	94.7	93.3	93.2	95.5	95.5	[96]	95.5	95.4
0.4	0	3.5^	2.5	2.6	2.4	2.6	2.6	[95]	2.6	2.6	95.7	93.8	91.9	91.7	94.3	94.5	[95]	94.4	94.3
0.6	0	3.8^	2.5	2.6	2.5	2.5	2.5	[93]	2.5	2.4	95.2	93.0	90.5	90.3	92.9	93.3	[93]	93.0	92.8
0.8	0	4.6^	2.4	2.4	2.3*	2.5	2.5	[90]	2.6	2.5	94.5	91.6	88.4	88.1	90.8	91.6	[91]	90.9	90.7
0	0.2	1.6*	2.4	2.4	2.3*	2.5	2.5	[97]	2.6	2.5	97.9	98.0	97.7	97.7	98.7	98.7	[97]	98.7	98.7
0.2	0.2	2.0*	2.5	2.5	2.4	2.6	2.5	[96]	2.6	2.5	97.6	97.6	96.9	96.8	98.1	98.2	[96]	98.2	98.1
0.4	0.2	2.3*	2.4	2.5	2.4	2.4	2.4	[95]	2.5	2.4	96.9	96.7	95.8	95.7	97.1	97.2	[95]	97.1	97.1
0.6	0.2	2.8^	2.5	2.5	2.4	2.5	2.5	[93]	2.6	2.5	96.3	95.8	94.3	94.2	95.9	96.2	[93]	95.9	95.8
0.8	0.2	3.4^	2.4	2.4	2.3*	2.5	2.4	[90]	2.6	2.5	95.7	95.0	92.5	92.4	94.2	95.0	[91]	94.3	94.1
0.4	0.4	1.3*	2.4	2.4	2.3*	2.4	2.4	[95]	2.5	2.4	98.2	98.9	98.6	98.6	99.1	99.2	[95]	99.2	99.1
0.6	0.4	1.7*	2.5	2.6	2.5	2.5	2.6	[93]	2.6	2.5	97.6	98.4	97.6	97.6	98.4	98.6	[93]	98.4	98.3
0	0.5	0.3*	2.4	2.3*	2.3*	2.4	2.4	[91]	2.6	2.4	99.4	99.8	99.9	99.9	99.9	100.0	[91]	99.9	99.9
0.8	0.5	1.9*	2.6	2.6	2.5	2.7^	2.6	[90]	2.8^	2.6	97.7	98.8	97.9	97.9	98.5	98.8	[91]	98.5	98.4
0.6	0.6	0.8*	2.6	2.5	2.5	2.5	2.6	[93]	2.7^	2.5	98.9	99.7	99.6	99.6	99.7	99.8	[93]	99.8	99.7
0.8	0.8	0.3*	2.8^	2.5	2.7^	2.7^	2.8^	[90]	2.9^	2.6	99.6	100.0	100.0	100.0	100.0	100.0	[91]	100.0	100.0

A.2. Chapter 2 Tables: Comparing a Method for Independent Data to a Method for Matched Data When Assessing Non-Inferiority via Risk Difference in Propensity-Score Matched Studies When There Are More Control Subjects Than Experimental Subjects

Table 2.1A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		*Max # controls per treated*	*Mean # controls per treated*	*St. dev. of # controls per treated*	*Mean % of treated subjects matched*	*Type I Error (%)*			*Bias of Est. RD for treated*		*Mean Est. SE of RD / St. dev. of Empirical Est. RD*	
Outcome	*Treatment*					*FM*	*AFM*	*AFMCC*	*FM*	*AFM*	*FM*	*AFM*
weak	weak	5	5.00	0.04	100.0	2.1	2.4	2.3	-0.000	-0.000	1.02	1.00
weak	strong	5	4.87	0.56	99.3	1.6	2.5	2.5	-0.004	-0.000	1.01	0.99
mixed	mixed	5	4.99	0.13	99.9	1.7	2.0	2.0	-0.000	0.000	1.06	1.05
strong	weak	5	5.00	0.04	100.0	1.2	2.4	2.2	0.000	0.000	1.13	1.01
strong	strong	5	4.87	0.56	99.3	0.5	2.3	2.1	-0.013	-0.001	1.09	1.00
weak	weak	2	2.00	0.00	100.0	2.2	2.5	2.4	0.000	0.000	1.02	1.00
weak	strong	2	1.99	0.07	99.5	2.0	2.4	2.3	-0.000	0.000	1.02	1.00
mixed	mixed	2	2.00	0.01	99.9	1.9	2.2	2.1	0.000	0.000	1.05	1.04
strong	weak	2	2.00	0.00	100.0	1.3	2.3	2.2	0.000	0.000	1.13	1.01
strong	strong	2	1.99	0.07	99.5	1.2	2.5	2.4	-0.002	-0.000	1.11	0.99
weak	weak	1	1.00	0.00	100.0	2.4	2.7	2.5	-0.000	-0.000	1.02	0.99

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Type I Error (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	strong	1	1.00	0.00	99.6	2.2	2.4	2.3	0.000	0.000	1.02	1.00
mixed	mixed	1	1.00	0.00	99.9	2.1	2.2	2.0	-0.000	-0.000	1.03	1.02
strong	weak	1	1.00	0.00	100.0	1.4	2.5	2.3	0.000	0.000	1.12	1.01
strong	strong	1	1.00	0.00	99.6	1.3	2.4	2.2	-0.000	-0.000	1.12	1.00

Table 2.1B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Power (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	Treatment					FM	AFM	AFMCC	FM	AFM	FM	AFM
weak	weak	5	5.00	0.04	100.0	91.7	92.5	92.4	-0.000	-0.000	1.14	1.11
weak	strong	5	4.87	0.56	99.3	88.8	91.9	91.8	-0.004	-0.000	1.12	1.09
mixed	mixed	5	4.99	0.14	99.9	91.9	92.9	92.8	-0.000	0.001	1.15	1.13
strong	weak	5	5.00	0.04	100.0	93.6	96.1	95.9	-0.000	0.000	1.25	1.14
strong	strong	5	4.87	0.56	99.3	84.2	95.4	95.1	-0.013	-0.001	1.18	1.11
weak	weak	2	2.00	0.00	100.0	85.9	87.2	86.8	-0.000	-0.000	1.08	1.05
weak	strong	2	1.99	0.07	99.5	85.3	86.8	86.4	-0.001	0.000	1.06	1.04
mixed	mixed	2	2.00	0.01	99.9	86.5	87.4	86.9	0.000	0.000	1.09	1.07
strong	weak	2	2.00	0.00	100.0	88.7	92.5	92.0	-0.000	0.000	1.18	1.07
strong	strong	2	1.99	0.07	99.5	87.2	92.6	92.1	-0.002	-0.000	1.19	1.08
weak	weak	1	1.00	0.00	100.0	79.0	80.0	79.1	0.001	0.001	1.03	1.02
weak	strong	1	1.00	0.00	99.6	77.7	78.5	77.4	-0.000	-0.000	1.02	1.02
mixed	mixed	1	1.00	0.00	99.9	78.3	78.6	77.7	-0.000	-0.000	1.03	1.03
strong	weak	1	1.00	0.00	100.0	81.0	86.0	85.2	0.000	0.000	1.13	1.03
strong	strong	1	1.00	0.00	99.6	80.7	85.7	85.0	0.000	0.000	1.12	1.02

Table 2.1C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null: Coverage of 95% CI		Under null: Length FM CI / length AFM CI	Under alternative: Coverage of 95% CI		Under alternative: Length FM CI / length AFM CI
Outcome	***Treatment***				***FM***	***AFM***		***FM***	***AFM***	
weak	weak	5	5.00	100.0	0.95	0.95	1.02	0.97	0.97	1.03
weak	strong	5	4.87	99.3	0.95	0.95	1.02	0.97	0.96	1.02
mixed	mixed	5	4.99	99.9	0.96	0.96	1.01	0.97	0.97	1.02
strong	weak	5	5.00	100.0	0.97	0.95	1.12	0.98	0.97	1.10
strong	strong	5	4.87	99.3	0.95	0.95	1.11	0.96	0.97	1.08
weak	weak	2	2.00	100.0	0.95	0.95	1.02	0.97	0.96	1.03
weak	strong	2	1.99	99.5	0.96	0.95	1.02	0.96	0.96	1.02
mixed	mixed	2	2.00	99.9	0.96	0.96	1.01	0.97	0.97	1.02
strong	weak	2	2.00	100.0	0.97	0.96	1.12	0.98	0.96	1.11
strong	strong	2	1.99	99.5	0.97	0.95	1.12	0.98	0.97	1.11
weak	weak	1	1.00	100.0	0.95	0.95	1.02	0.95	0.95	1.01
weak	strong	1	1.00	99.6	0.96	0.95	1.02	0.96	0.95	1.01
mixed	mixed	1	1.00	99.9	0.96	0.96	1.01	0.96	0.96	1.00
strong	weak	1	1.00	100.0	0.97	0.95	1.12	0.97	0.95	1.10
strong	strong	1	1.00	99.6	0.97	0.95	1.12	0.97	0.96	1.10

Table 2.2A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated normal & bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Type I Error (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	Treatment					FM	AFM	AFMCC	FM	AFM	FM	AFM
weak	weak	5	5.00	0.04	100.0	2.1	2.4	2.4	0.000	0.000	1.02	0.99
weak	strong	5	4.87	0.57	99.3	1.5	2.5	2.4	-0.004	0.000	1.01	0.99
mixed	mixed	5	4.99	0.13	99.9	1.6	1.9	1.8	-0.000	0.000	1.06	1.04
strong	weak	5	5.00	0.04	100.0	1.3	2.1	2.0	-0.001	-0.000	1.13	1.01
strong	strong	5	4.87	0.57	99.3	0.4	1.9	1.8	-0.013	-0.001	1.10	1.01
weak	weak	2	2.00	0.00	100.0	2.0	2.3	2.2	-0.000	-0.000	1.03	1.01
weak	strong	2	1.99	0.07	99.4	2.1	2.5	2.4	-0.001	-0.000	1.02	1.00
mixed	mixed	2	2.00	0.01	99.9	2.0	2.3	2.2	-0.000	-0.000	1.05	1.03
strong	weak	2	2.00	0.00	100.0	1.1	2.5	2.3	-0.000	0.000	1.14	1.02
strong	strong	2	1.99	0.07	99.4	1.0	2.3	2.2	-0.003	-0.001	1.12	1.00
weak	weak	1	1.00	0.00	100.0	2.1	2.3	2.1	-0.000	-0.000	1.02	1.00
weak	strong	1	1.00	0.00	99.5	2.5	2.7	2.5	0.001	0.001	1.01	0.99
mixed	mixed	1	1.00	0.00	99.9	2.1	2.2	2.1	-0.000	-0.000	1.04	1.02
strong	weak	1	1.00	0.00	100.0	1.4	2.5	2.3	-0.000	-0.000	1.12	1.00
strong	strong	1	1.00	0.00	99.6	1.4	2.7	2.4	-0.000	-0.000	1.13	1.00

Table 2.2B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated normal & bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Power (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	5.00	0.04	100.0	91.5	92.3	92.1	-0.000	0.000	1.12	1.09
weak	strong	5	4.87	0.56	99.3	88.9	92.1	91.9	-0.005	-0.000	1.13	1.10
mixed	mixed	5	4.99	0.13	99.9	92.2	93.0	92.8	-0.001	0.000	1.16	1.14
strong	weak	5	5.00	0.04	100.0	93.6	96.1	96.0	-0.000	0.000	1.25	1.14
strong	strong	5	4.87	0.56	99.3	84.6	95.1	94.9	-0.013	-0.001	1.18	1.11
weak	weak	2	2.00	0.00	100.0	86.4	87.5	87.0	-0.000	-0.000	1.08	1.06
weak	strong	2	1.99	0.07	99.4	86.1	87.4	86.9	-0.001	0.000	1.09	1.06
mixed	mixed	2	2.00	0.01	99.9	86.9	88.0	87.6	-0.000	0.000	1.11	1.09
strong	weak	2	2.00	0.00	100.0	88.5	92.5	92.1	-0.000	-0.000	1.18	1.06
strong	strong	2	1.99	0.07	99.4	87.0	92.5	92.1	-0.002	-0.000	1.18	1.07
weak	weak	1	1.00	0.00	100.0	79.0	79.8	78.9	-0.000	-0.000	1.04	1.03
weak	strong	1	1.00	0.00	99.5	78.1	78.7	77.9	-0.000	-0.000	1.02	1.01
mixed	mixed	1	1.00	0.00	99.9	78.6	79.3	78.3	-0.000	-0.000	1.04	1.03
strong	weak	1	1.00	0.00	100.0	80.4	85.8	85.1	0.000	0.000	1.13	1.03
strong	strong	1	1.00	0.00	99.6	79.9	85.3	84.6	-0.000	-0.000	1.11	1.01

Table 2.2C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated normal & bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null: Coverage of 95% CI		Under null: Length FM CI / length AFM CI	Under alternative: Coverage of 95% CI		Under alternative: Length FM CI / length AFM CI
Outcome	***Treatment***				***FM***	***AFM***		***FM***	***AFM***	
weak	weak	5	5.00	100.0	0.95	0.95	1.02	0.97	0.97	1.02
weak	strong	5	4.87	99.3	0.95	0.95	1.02	0.97	0.97	1.02
mixed	mixed	5	4.99	99.9	0.96	0.96	1.02	0.98	0.98	1.02
strong	weak	5	5.00	100.0	0.97	0.95	1.12	0.99	0.98	1.10
strong	strong	5	4.87	99.3	0.95	0.95	1.11	0.96	0.97	1.08
weak	weak	2	2.00	100.0	0.96	0.95	1.02	0.97	0.96	1.03
weak	strong	2	1.99	99.4	0.95	0.95	1.02	0.97	0.96	1.02
mixed	mixed	2	2.00	99.9	0.96	0.96	1.02	0.97	0.97	1.02
strong	weak	2	2.00	100.0	0.98	0.95	1.12	0.98	0.96	1.11
strong	strong	2	1.99	99.4	0.97	0.95	1.12	0.98	0.96	1.11
weak	weak	1	1.00	100.0	0.96	0.95	1.02	0.96	0.96	1.01
weak	strong	1	1.00	99.5	0.95	0.95	1.02	0.95	0.95	1.01
mixed	mixed	1	1.00	99.9	0.96	0.96	1.02	0.96	0.96	1.01
strong	weak	1	1.00	100.0	0.97	0.95	1.12	0.97	0.96	1.10
strong	strong	1	1.00	99.6	0.97	0.95	1.12	0.97	0.95	1.10

Table 2.3A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Type I Error (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	Treatment					FM	AFM	AFMCC	FM	AFM	FM	AFM
weak	weak	5	4.99	0.07	100.0	2.0	2.5	2.4	-0.000	0.000	1.05	1.01
weak	strong	5	4.64	0.96	97.5	1.0	2.6	2.5	-0.011	0.000	1.02	0.99
mixed	mixed	5	4.95	0.31	99.7	1.4	1.8	1.8	-0.002	-0.000	1.08	1.06
strong	weak	5	4.99	0.07	100.0	1.0	2.3	2.1	-0.001	0.000	1.19	1.02
strong	strong	5	4.64	0.96	97.5	0.1	1.9	1.8	-0.032	-0.003	1.08	0.99
weak	weak	2	2.00	0.01	100.0	1.9	2.3	2.2	-0.000	-0.000	1.04	1.00
weak	strong	2	1.97	0.15	98.1	1.6	2.3	2.2	-0.002	0.000	1.03	1.01
mixed	mixed	2	2.00	0.03	99.8	1.8	2.0	1.9	-0.000	-0.000	1.06	1.05
strong	weak	2	2.00	0.01	100.0	0.9	2.3	2.1	-0.000	0.000	1.20	1.03
strong	strong	2	1.97	0.15	98.1	0.6	2.4	2.2	-0.007	-0.001	1.14	1.01
weak	weak	1	1.00	0.00	100.0	1.9	2.2	2.1	-0.000	-0.000	1.05	1.01
weak	strong	1	1.00	0.00	98.6	2.3	2.5	2.3	0.001	0.001	1.03	1.01
mixed	mixed	1	1.00	0.00	99.8	2.0	2.1	1.9	0.000	0.000	1.06	1.04
strong	weak	1	1.00	0.00	100.0	1.0	2.5	2.3	0.000	0.000	1.18	1.01
strong	strong	1	1.00	0.00	98.6	1.2	2.6	2.3	-0.001	-0.001	1.15	1.00

Table 2.3B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Power (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	4.99	0.07	100.0	91.4	92.9	92.7	-0.001	-0.000	1.14	1.10
weak	strong	5	4.64	0.96	97.5	82.9	91.3	91.1	-0.012	-0.001	1.12	1.07
mixed	mixed	5	4.95	0.31	99.7	91.9	93.3	93.2	-0.002	0.000	1.18	1.16
strong	weak	5	4.99	0.07	100.0	94.1	97.4	97.3	-0.001	-0.000	1.31	1.15
strong	strong	5	4.64	0.96	97.5	64.9	95.2	94.9	-0.031	-0.002	1.17	1.10
weak	weak	2	2.00	0.01	100.0	87.1	88.7	88.2	0.000	0.000	1.10	1.07
weak	strong	2	1.97	0.15	98.1	84.3	87.2	86.8	-0.003	-0.000	1.08	1.05
mixed	mixed	2	2.00	0.03	99.8	87.4	88.5	88.0	-0.000	0.000	1.13	1.10
strong	weak	2	2.00	0.01	100.0	89.6	94.3	94.0	0.000	0.000	1.25	1.08
strong	strong	2	1.97	0.15	98.1	84.3	93.3	92.8	-0.007	-0.001	1.19	1.07
weak	weak	1	1.00	0.00	100.0	78.8	80.2	79.4	0.000	0.000	1.05	1.02
weak	strong	1	1.00	0.00	98.6	78.8	79.8	78.8	0.000	0.000	1.03	1.02
mixed	mixed	1	1.00	0.00	99.8	79.1	79.6	78.9	0.000	0.000	1.05	1.04
strong	weak	1	1.00	0.00	100.0	81.9	89.3	88.5	-0.000	-0.000	1.20	1.04
strong	strong	1	1.00	0.00	98.6	81.1	87.3	86.4	-0.000	-0.000	1.15	1.02

Table 2.3C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null			Under alternative		
					Coverage of 95% CI		Length FM CI / length AFM CI	Coverage of 95% CI		Length FM CI / length AFM CI
Outcome	Treatment				FM	AFM		FM	AFM	
weak	weak	5	4.99	100.0	0.96	0.95	1.04	0.97	0.97	1.04
weak	strong	5	4.64	97.5	0.94	0.95	1.02	0.96	0.96	1.02
mixed	mixed	5	4.95	99.7	0.97	0.96	1.02	0.98	0.98	1.02
strong	weak	5	4.99	100.0	0.98	0.95	1.17	0.99	0.97	1.14
strong	strong	5	4.64	97.5	0.84	0.95	1.10	0.86	0.97	1.08
weak	weak	2	2.00	100.0	0.96	0.95	1.04	0.97	0.96	1.04
weak	strong	2	1.97	98.1	0.96	0.95	1.02	0.97	0.96	1.03
mixed	mixed	2	2.00	99.8	0.96	0.96	1.02	0.97	0.97	1.02
strong	weak	2	2.00	100.0	0.98	0.96	1.17	0.99	0.97	1.16
strong	strong	2	1.97	98.1	0.97	0.95	1.14	0.98	0.97	1.12
weak	weak	1	1.00	100.0	0.96	0.95	1.04	0.96	0.95	1.02
weak	strong	1	1.00	98.6	0.96	0.95	1.02	0.96	0.95	1.01
mixed	mixed	1	1.00	99.8	0.96	0.96	1.02	0.96	0.96	1.01
strong	weak	1	1.00	100.0	0.98	0.95	1.17	0.98	0.96	1.15
strong	strong	1	1.00	98.6	0.98	0.95	1.15	0.98	0.96	1.13

Table 2.4A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated normal covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Type I Error (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	4.98	0.16	99.9	1.6	2.5	2.4	-0.001	0.000	1.07	1.00
weak	strong	5	4.08	1.45	88.7	0.4	3.3	3.2	-0.021	0.001	1.00	0.96
mixed	mixed	5	4.80	0.71	98.8	0.4	1.9	1.7	-0.013	-0.001	1.09	1.04
strong	weak	5	4.98	0.16	99.9	0.3	2.1	1.8	-0.003	-0.000	1.31	1.01
strong	strong	5	4.08	1.45	88.7	0.0	2.6	2.4	-0.049	-0.002	1.03	0.95
weak	weak	2	2.00	0.02	99.9	1.8	2.6	2.5	-0.000	-0.000	1.06	1.00
weak	strong	2	1.90	0.30	90.2	1.4	2.9	2.8	-0.004	0.002	1.03	1.00
mixed	mixed	2	1.99	0.09	99.1	1.0	1.9	1.8	-0.002	-0.000	1.12	1.05
strong	weak	2	2.00	0.02	99.9	0.5	2.4	2.1	-0.000	0.000	1.32	1.01
strong	strong	2	1.90	0.30	90.2	0.2	2.3	2.1	-0.018	-0.002	1.11	0.99
weak	weak	1	1.00	0.00	99.9	2.1	2.6	2.6	0.000	0.000	1.06	1.00
weak	strong	1	1.00	0.00	92.4	2.4	2.7	2.5	0.002	0.002	1.03	1.00
mixed	mixed	1	1.00	0.00	99.3	1.5	2.1	2.0	-0.000	-0.000	1.11	1.04
strong	weak	1	1.00	0.00	99.9	0.5	2.7	2.4	0.000	0.000	1.31	1.00
strong	strong	1	1.00	0.00	92.4	1.1	2.3	2.1	-0.001	-0.001	1.14	1.00

Table 2.4B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated normal covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Power (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	4.98	0.16	99.9	91.3	93.9	93.7	-0.002	-0.000	1.17	1.11
weak	strong	5	4.08	1.45	88.6	67.6	87.8	87.5	-0.025	-0.003	1.11	1.03
mixed	mixed	5	4.80	0.71	98.9	84.5	94.5	94.3	-0.013	-0.001	1.19	1.14
strong	weak	5	4.98	0.16	99.9	95.4	99.1	99.0	-0.002	0.000	1.44	1.18
strong	strong	5	4.08	1.45	88.6	33.5	92.3	91.9	-0.055	-0.008	1.18	1.05
weak	weak	2	2.00	0.02	99.9	87.3	89.9	89.5	-0.000	-0.000	1.12	1.06
weak	strong	2	1.90	0.29	90.2	78.1	84.3	83.8	-0.009	-0.002	1.08	1.03
mixed	mixed	2	1.99	0.10	99.1	87.0	90.9	90.4	-0.002	0.000	1.17	1.10
strong	weak	2	2.00	0.02	99.9	92.0	97.9	97.7	-0.000	-0.000	1.40	1.10
strong	strong	2	1.90	0.29	90.2	72.8	91.4	90.8	-0.022	-0.006	1.20	1.08
weak	weak	1	1.00	0.00	99.9	79.2	82.2	81.4	-0.000	-0.000	1.07	1.02
weak	strong	1	1.00	0.00	92.4	77.3	78.2	77.1	-0.001	-0.001	1.04	1.03
mixed	mixed	1	1.00	0.00	99.3	79.9	83.0	82.2	-0.001	-0.001	1.10	1.04
strong	weak	1	1.00	0.00	99.9	84.3	93.9	93.5	-0.000	-0.000	1.32	1.03
strong	strong	1	1.00	0.00	92.4	82.3	87.5	86.7	-0.003	-0.003	1.16	1.04

Table 2.4C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated normal covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null: Coverage of 95% CI		Under null: Length FM CI / length AFM CI	Under alternative: Coverage of 95% CI		Under alternative: Length FM CI / length AFM CI
Outcome	*Treatment*				*FM*	*AFM*		*FM*	*AFM*	
weak	weak	5	4.98	99.9	0.96	0.95	1.07	0.98	0.97	1.06
weak	strong	5	4.08	88.7	0.91	0.94	1.01	0.90	0.96	1.01
mixed	mixed	5	4.80	98.8	0.95	0.96	1.06	0.96	0.98	1.05
strong	weak	5	4.98	99.9	0.99	0.95	1.30	0.99	0.98	1.24
strong	strong	5	4.08	88.7	0.61	0.94	1.04	0.50	0.95	1.03
weak	weak	2	2.00	99.9	0.96	0.95	1.07	0.97	0.96	1.06
weak	strong	2	1.90	90.2	0.96	0.95	1.02	0.96	0.96	1.02
mixed	mixed	2	1.99	99.1	0.97	0.96	1.07	0.98	0.97	1.06
strong	weak	2	2.00	99.9	0.99	0.95	1.31	0.99	0.97	1.28
strong	strong	2	1.90	90.2	0.94	0.95	1.10	0.93	0.96	1.08
weak	weak	1	1.00	99.9	0.96	0.95	1.07	0.96	0.95	1.05
weak	strong	1	1.00	92.4	0.96	0.95	1.03	0.96	0.96	1.02
mixed	mixed	1	1.00	99.3	0.97	0.96	1.07	0.97	0.96	1.06
strong	weak	1	1.00	99.9	0.99	0.95	1.31	0.99	0.96	1.29
strong	strong	1	1.00	92.4	0.98	0.95	1.14	0.98	0.96	1.11

Table 2.5A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		*Max # controls per treated*	*Mean # controls per treated*	*St. dev. of # controls per treated*	*Mean % of treated subjects matched*	*Type I Error (%)*			*Bias of Est. RD for treated*		*Mean Est. SE of RD / St. dev. of Empirical Est. RD*	
Outcome	*Treatment*					*FM*	*AFM*	*AFMCC*	*FM*	*AFM*	*FM*	*AFM*
weak	weak	5	5.00	0.00	100.0	2.2	2.4	2.3	0.000	0.000	1.00	0.99
weak	strong	5	5.00	0.00	100.0	2.3	2.4	2.4	0.000	0.000	1.01	1.00
mixed	mixed	5	5.00	0.00	100.0	2.2	2.7	2.5	-0.000	-0.000	1.02	0.99
strong	weak	5	5.00	0.00	100.0	1.7	2.3	2.2	0.000	0.000	1.06	1.00
strong	strong	5	5.00	0.00	100.0	1.9	2.5	2.4	-0.000	-0.000	1.05	1.00
weak	weak	2	2.00	0.00	100.0	2.3	2.6	2.4	-0.000	-0.000	1.00	0.99
weak	strong	2	2.00	0.00	100.0	2.0	2.2	2.1	0.000	0.000	1.02	1.01
mixed	mixed	2	2.00	0.00	100.0	2.0	2.5	2.3	0.000	0.000	1.04	1.02
strong	weak	2	2.00	0.00	100.0	1.5	2.2	2.1	-0.000	-0.000	1.07	1.00
strong	strong	2	2.00	0.00	100.0	1.8	2.4	2.3	0.000	0.000	1.05	1.00
weak	weak	1	1.00	0.00	100.0	2.3	2.3	2.2	-0.000	-0.000	1.01	1.00
weak	strong	1	1.00	0.00	100.0	2.5	2.6	2.4	0.001	0.001	1.00	0.99
mixed	mixed	1	1.00	0.00	100.0	2.1	2.5	2.3	-0.000	-0.000	1.03	1.00
strong	weak	1	1.00	0.00	100.0	1.9	2.7	2.5	0.000	0.000	1.06	1.00
strong	strong	1	1.00	0.00	100.0	2.1	2.7	2.6	-0.000	-0.000	1.04	0.99

Table 2.5B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Power (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	Treatment					FM	AFM	AFMCC	FM	AFM	FM	AFM
weak	weak	5	5.00	0.00	100.0	91.3	91.6	91.5	-0.000	-0.000	1.11	1.09
weak	strong	5	5.00	0.00	100.0	91.1	91.5	91.3	-0.000	-0.000	1.10	1.09
mixed	mixed	5	5.00	0.00	100.0	92.1	93.0	92.8	0.000	0.000	1.13	1.10
strong	weak	5	5.00	0.00	100.0	92.2	94.0	93.8	0.000	0.000	1.17	1.11
strong	strong	5	5.00	0.00	100.0	92.2	93.6	93.3	-0.000	-0.000	1.16	1.11
weak	weak	2	2.00	0.00	100.0	86.4	87.0	86.6	0.000	0.000	1.08	1.06
weak	strong	2	2.00	0.00	100.0	86.3	87.0	86.4	-0.000	-0.000	1.07	1.06
mixed	mixed	2	2.00	0.00	100.0	86.7	87.9	87.4	0.000	0.000	1.09	1.05
strong	weak	2	2.00	0.00	100.0	87.5	89.8	89.3	-0.000	-0.000	1.12	1.06
strong	strong	2	2.00	0.00	100.0	87.1	89.2	88.6	-0.000	-0.000	1.12	1.06
weak	weak	1	1.00	0.00	100.0	77.6	77.7	76.8	-0.001	-0.001	1.01	1.01
weak	strong	1	1.00	0.00	100.0	77.9	78.3	77.4	-0.000	-0.000	1.01	1.01
mixed	mixed	1	1.00	0.00	100.0	78.4	79.5	78.5	0.000	0.000	1.04	1.02
strong	weak	1	1.00	0.00	100.0	79.5	82.1	81.2	0.000	0.000	1.07	1.02
strong	strong	1	1.00	0.00	100.0	79.0	81.2	80.5	0.000	0.000	1.05	1.01

Table 2.5C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = independent bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null: Coverage of 95% CI		Under null: Length FM CI / length AFM CI	Under alternative: Coverage of 95% CI		Under alternative: Length FM CI / length AFM CI
Outcome	Treatment				FM	AFM		FM	AFM	
weak	weak	5	5.00	100.0	0.95	0.95	1.01	0.97	0.97	1.02
weak	strong	5	5.00	100.0	0.95	0.95	1.01	0.97	0.97	1.01
mixed	mixed	5	5.00	100.0	0.95	0.95	1.03	0.97	0.97	1.03
strong	weak	5	5.00	100.0	0.96	0.95	1.07	0.98	0.97	1.06
strong	strong	5	5.00	100.0	0.96	0.95	1.06	0.98	0.97	1.05
weak	weak	2	2.00	100.0	0.95	0.95	1.01	0.96	0.96	1.02
weak	strong	2	2.00	100.0	0.95	0.95	1.01	0.96	0.96	1.01
mixed	mixed	2	2.00	100.0	0.96	0.96	1.03	0.97	0.96	1.03
strong	weak	2	2.00	100.0	0.97	0.95	1.07	0.97	0.96	1.06
strong	strong	2	2.00	100.0	0.96	0.95	1.05	0.97	0.96	1.05
weak	weak	1	1.00	100.0	0.95	0.95	1.01	0.95	0.95	1.00
weak	strong	1	1.00	100.0	0.95	0.95	1.01	0.95	0.95	1.00
mixed	mixed	1	1.00	100.0	0.96	0.95	1.03	0.96	0.96	1.02
strong	weak	1	1.00	100.0	0.96	0.95	1.06	0.96	0.96	1.05
strong	strong	1	1.00	100.0	0.96	0.95	1.05	0.96	0.95	1.04

Table 2.6A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		*Max # controls per treated*	*Mean # controls per treated*	*St. dev. of # controls per treated*	*Mean % of treated subjects matched*	*Type I Error (%)*			*Bias of Est. RD for treated*		*Mean Est. SE of RD / St. dev. of Empirical Est. RD*	
Outcome	*Treatment*					*FM*	*AFM*	*AFMCC*	*FM*	*AFM*	*FM*	*AFM*
weak	weak	5	5.00	0.00	100.0	2.2	2.4	2.3	-0.000	-0.000	1.03	1.01
weak	strong	5	5.00	0.00	100.0	2.4	2.5	2.4	0.000	0.000	1.02	1.01
mixed	mixed	5	5.00	0.00	100.0	2.1	2.6	2.5	0.000	0.000	1.04	1.00
strong	weak	5	5.00	0.00	100.0	1.5	2.4	2.3	0.000	0.000	1.09	1.00
strong	strong	5	5.00	0.00	100.0	2.0	2.6	2.5	0.000	0.000	1.05	0.99
weak	weak	2	2.00	0.00	100.0	2.2	2.5	2.4	-0.000	-0.000	1.02	1.00
weak	strong	2	2.00	0.00	100.0	2.2	2.3	2.2	-0.000	-0.000	1.02	1.01
mixed	mixed	2	2.00	0.00	100.0	2.2	2.7	2.6	-0.000	-0.000	1.02	0.99
strong	weak	2	2.00	0.00	100.0	1.5	2.3	2.2	-0.000	-0.000	1.09	1.00
strong	strong	2	2.00	0.00	100.0	2.0	2.6	2.5	0.000	0.000	1.06	1.01
weak	weak	1	1.00	0.00	100.0	2.3	2.4	2.2	0.000	0.000	1.01	1.00
weak	strong	1	1.00	0.00	100.0	2.3	2.4	2.3	-0.000	-0.000	1.01	1.00
mixed	mixed	1	1.00	0.00	100.0	2.1	2.4	2.3	-0.000	-0.000	1.03	0.99
strong	weak	1	1.00	0.00	100.0	1.6	2.6	2.3	-0.000	-0.000	1.09	1.00
strong	strong	1	1.00	0.00	100.0	2.0	2.6	2.4	0.000	0.000	1.05	1.00

Table 2.6B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Power (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	Treatment					FM	AFM	AFMCC	FM	AFM	FM	AFM
weak	weak	5	5.00	0.00	100.0	91.0	91.6	91.5	-0.000	-0.000	1.11	1.09
weak	strong	5	5.00	0.00	100.0	91.4	91.9	91.7	0.000	0.000	1.10	1.09
mixed	mixed	5	5.00	0.00	100.0	91.8	92.9	92.7	-0.000	-0.000	1.14	1.10
strong	weak	5	5.00	0.00	100.0	93.1	94.8	94.6	0.000	0.000	1.19	1.11
strong	strong	5	5.00	0.00	100.0	92.0	93.4	93.2	0.000	0.000	1.15	1.10
weak	weak	2	2.00	0.00	100.0	87.0	87.7	87.2	0.000	0.000	1.08	1.06
weak	strong	2	2.00	0.00	100.0	85.8	86.6	86.1	0.000	0.000	1.06	1.04
mixed	mixed	2	2.00	0.00	100.0	86.5	87.9	87.5	-0.000	-0.000	1.10	1.06
strong	weak	2	2.00	0.00	100.0	87.2	90.3	89.9	-0.000	-0.000	1.13	1.05
strong	strong	2	2.00	0.00	100.0	86.5	88.3	87.8	-0.000	-0.000	1.09	1.05
weak	weak	1	1.00	0.00	100.0	78.8	79.0	78.2	0.000	0.000	1.02	1.02
weak	strong	1	1.00	0.00	100.0	78.0	78.1	77.3	0.000	0.000	1.01	1.01
mixed	mixed	1	1.00	0.00	100.0	78.7	80.2	79.4	-0.000	-0.000	1.04	1.02
strong	weak	1	1.00	0.00	100.0	79.5	83.5	82.7	0.000	0.000	1.09	1.02
strong	strong	1	1.00	0.00	100.0	78.8	80.8	79.9	0.000	0.000	1.05	1.02

Table 2.6C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 20000

					Under null			Under alternative		
Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Coverage of 95% CI		Length FM CI / length AFM CI	Coverage of 95% CI		Length FM CI / length AFM CI
Outcome	Treatment				FM	AFM		FM	AFM	
weak	weak	5	5.00	100.0	0.95	0.95	1.01	0.97	0.97	1.02
weak	strong	5	5.00	100.0	0.95	0.95	1.01	0.97	0.97	1.02
mixed	mixed	5	5.00	100.0	0.96	0.95	1.04	0.97	0.97	1.03
strong	weak	5	5.00	100.0	0.97	0.95	1.10	0.98	0.97	1.07
strong	strong	5	5.00	100.0	0.96	0.95	1.05	0.98	0.97	1.04
weak	weak	2	2.00	100.0	0.95	0.95	1.01	0.96	0.96	1.02
weak	strong	2	2.00	100.0	0.95	0.95	1.01	0.96	0.96	1.02
mixed	mixed	2	2.00	100.0	0.96	0.95	1.04	0.97	0.96	1.04
strong	weak	2	2.00	100.0	0.97	0.95	1.09	0.97	0.96	1.08
strong	strong	2	2.00	100.0	0.96	0.95	1.05	0.97	0.96	1.05
weak	weak	1	1.00	100.0	0.95	0.95	1.01	0.96	0.95	1.00
weak	strong	1	1.00	100.0	0.95	0.95	1.01	0.95	0.95	1.00
mixed	mixed	1	1.00	100.0	0.96	0.95	1.04	0.96	0.95	1.02
strong	weak	1	1.00	100.0	0.97	0.95	1.09	0.97	0.96	1.07
strong	strong	1	1.00	100.0	0.96	0.95	1.05	0.96	0.95	1.03

Table 2.7A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.1 based on raw propensity score, total sample size (pre-match) = 20000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Type I Error (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	5.00	0.00	100.0	2.0	2.4	2.3	0.000	0.000	1.02	1.00
weak	strong	5	4.92	0.45	99.5	1.8	2.6	2.5	-0.003	-0.000	1.01	0.99
mixed	mixed	5	5.00	0.03	100.0	1.7	1.8	1.8	-0.000	-0.000	1.05	1.04
strong	weak	5	5.00	0.00	100.0	1.3	2.4	2.2	0.000	0.000	1.13	1.01
strong	strong	5	4.92	0.45	99.5	0.5	1.8	1.7	-0.011	-0.003	1.10	1.01
weak	weak	2	2.00	0.00	100.0	2.2	2.5	2.4	0.000	0.000	1.02	1.00
weak	strong	2	1.99	0.05	99.6	2.0	2.4	2.3	-0.001	-0.000	1.03	1.01
mixed	mixed	2	2.00	0.00	100.0	1.7	1.9	1.8	-0.000	-0.000	1.06	1.04
strong	weak	2	2.00	0.00	100.0	1.3	2.5	2.3	-0.000	-0.000	1.13	1.01
strong	strong	2	1.99	0.05	99.6	1.1	2.4	2.2	-0.002	-0.001	1.12	1.00
weak	weak	1	1.00	0.00	100.0	2.4	2.6	2.5	0.000	0.000	1.02	1.00
weak	strong	1	1.00	0.00	99.7	2.3	2.5	2.4	-0.000	-0.000	1.02	1.00
mixed	mixed	1	1.00	0.00	100.0	2.0	2.2	2.0	-0.000	-0.000	1.04	1.03
strong	weak	1	1.00	0.00	100.0	1.2	2.3	2.1	0.000	0.000	1.12	1.01
strong	strong	1	1.00	0.00	99.7	1.4	2.5	2.4	-0.000	-0.000	1.12	0.99

Table 2.7B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.1 based on raw propensity score, total sample size (pre-match) = 20000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Power (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	5.00	0.00	100.0	91.5	92.3	92.1	-0.000	-0.000	1.13	1.10
weak	strong	5	4.92	0.45	99.5	89.0	91.5	91.3	-0.004	-0.001	1.12	1.09
mixed	mixed	5	5.00	0.03	100.0	92.1	92.6	92.5	-0.000	-0.000	1.16	1.14
strong	weak	5	5.00	0.00	100.0	93.7	96.1	96.0	0.000	0.000	1.24	1.13
strong	strong	5	4.92	0.45	99.5	86.7	94.9	94.7	-0.011	-0.003	1.21	1.13
weak	weak	2	2.00	0.00	100.0	86.9	87.9	87.4	-0.000	-0.000	1.10	1.07
weak	strong	2	1.99	0.05	99.6	85.6	87.0	86.5	-0.001	-0.000	1.08	1.05
mixed	mixed	2	2.00	0.00	100.0	87.2	87.9	87.3	-0.000	-0.000	1.11	1.09
strong	weak	2	2.00	0.00	100.0	88.4	92.3	91.7	-0.000	-0.000	1.19	1.08
strong	strong	2	1.99	0.05	99.6	87.2	92.0	91.7	-0.002	-0.001	1.17	1.06
weak	weak	1	1.00	0.00	100.0	77.9	78.7	77.9	-0.001	-0.001	1.03	1.02
weak	strong	1	1.00	0.00	99.7	78.6	79.2	78.4	-0.000	-0.000	1.03	1.03
mixed	mixed	1	1.00	0.00	100.0	78.4	78.9	77.9	0.000	0.000	1.03	1.03
strong	weak	1	1.00	0.00	100.0	80.3	85.4	84.6	-0.000	-0.000	1.12	1.02
strong	strong	1	1.00	0.00	99.7	80.2	85.2	84.4	-0.000	-0.000	1.12	1.02

Table 2.7C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.1 based on raw propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null: Coverage of 95% CI		Under null: Length FM CI / length AFM CI	Under alternative: Coverage of 95% CI		Under alternative: Length FM CI / length AFM CI
Outcome	***Treatment***				***FM***	***AFM***		***FM***	***AFM***	
weak	weak	5	5.00	100.0	0.96	0.95	1.02	0.97	0.97	1.03
weak	strong	5	4.92	99.5	0.95	0.95	1.02	0.97	0.97	1.02
mixed	mixed	5	5.00	100.0	0.96	0.96	1.01	0.98	0.98	1.02
strong	weak	5	5.00	100.0	0.97	0.95	1.12	0.98	0.97	1.10
strong	strong	5	4.92	99.5	0.96	0.95	1.11	0.97	0.97	1.09
weak	weak	2	2.00	100.0	0.95	0.95	1.02	0.97	0.96	1.03
weak	strong	2	1.99	99.6	0.96	0.95	1.02	0.97	0.96	1.02
mixed	mixed	2	2.00	100.0	0.96	0.96	1.01	0.97	0.97	1.02
strong	weak	2	2.00	100.0	0.97	0.95	1.12	0.98	0.96	1.11
strong	strong	2	1.99	99.6	0.97	0.95	1.12	0.98	0.96	1.11
weak	weak	1	1.00	100.0	0.95	0.95	1.02	0.95	0.95	1.01
weak	strong	1	1.00	99.7	0.95	0.95	1.02	0.96	0.96	1.01
mixed	mixed	1	1.00	100.0	0.96	0.96	1.01	0.96	0.96	1.00
strong	weak	1	1.00	100.0	0.98	0.95	1.12	0.97	0.96	1.10
strong	strong	1	1.00	99.7	0.97	0.95	1.13	0.97	0.96	1.10

Table 2.8A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.5*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Type I Error (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	5.00	0.01	100.0	2.0	2.4	2.3	-0.000	-0.000	1.02	0.99
weak	strong	5	4.92	0.43	99.7	1.7	2.3	2.2	-0.004	-0.001	1.00	0.98
mixed	mixed	5	4.99	0.06	100.0	1.7	1.9	1.8	-0.000	-0.000	1.07	1.06
strong	weak	5	5.00	0.01	100.0	1.3	2.3	2.1	0.000	0.000	1.14	1.02
strong	strong	5	4.92	0.43	99.7	0.5	1.6	1.5	-0.012	-0.004	1.09	0.99
weak	weak	2	2.00	0.00	100.0	2.1	2.4	2.4	-0.000	-0.000	1.02	1.00
weak	strong	2	2.00	0.04	99.8	2.0	2.4	2.3	-0.001	-0.001	1.02	1.00
mixed	mixed	2	2.00	0.00	100.0	1.7	1.9	1.9	-0.001	-0.000	1.05	1.04
strong	weak	2	2.00	0.00	100.0	1.4	2.6	2.4	-0.000	-0.000	1.13	1.01
strong	strong	2	2.00	0.04	99.8	1.2	2.5	2.3	-0.002	-0.001	1.12	1.00
weak	weak	1	1.00	0.00	100.0	2.1	2.4	2.2	-0.001	-0.001	1.03	1.01
weak	strong	1	1.00	0.00	99.9	2.3	2.5	2.3	0.000	0.000	1.01	0.99
mixed	mixed	1	1.00	0.00	100.0	1.9	2.0	1.9	0.000	0.000	1.04	1.03
strong	weak	1	1.00	0.00	100.0	1.3	2.3	2.2	-0.000	-0.000	1.13	1.01
strong	strong	1	1.00	0.00	99.9	1.4	2.6	2.4	-0.000	-0.000	1.12	1.00

Table 2.8B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.5*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Power (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	**Treatment**					**FM**	**AFM**	**AFMCC**	**FM**	**AFM**	**FM**	**AFM**
weak	weak	5	5.00	0.01	100.0	91.5	92.3	92.0	-0.000	-0.000	1.14	1.11
weak	strong	5	4.92	0.43	99.7	89.6	91.7	91.6	-0.004	-0.001	1.12	1.09
mixed	mixed	5	4.99	0.06	100.0	92.1	92.8	92.6	-0.000	-0.000	1.15	1.13
strong	weak	5	5.00	0.01	100.0	93.1	95.7	95.4	-0.000	0.000	1.24	1.13
strong	strong	5	4.92	0.43	99.7	85.7	94.4	94.0	-0.011	-0.004	1.19	1.11
weak	weak	2	2.00	0.00	100.0	86.3	87.3	86.7	-0.000	-0.000	1.08	1.05
weak	strong	2	2.00	0.04	99.8	86.1	87.5	87.0	-0.001	-0.000	1.09	1.07
mixed	mixed	2	2.00	0.00	100.0	87.5	88.1	87.7	0.000	0.000	1.11	1.09
strong	weak	2	2.00	0.00	100.0	89.0	92.6	92.1	0.000	0.000	1.19	1.08
strong	strong	2	2.00	0.04	99.8	87.2	92.1	91.6	-0.002	-0.001	1.18	1.06
weak	weak	1	1.00	0.00	100.0	78.2	79.1	78.3	-0.000	-0.000	1.03	1.01
weak	strong	1	1.00	0.00	99.9	78.4	79.4	78.4	0.000	0.000	1.03	1.02
mixed	mixed	1	1.00	0.00	100.0	78.4	78.6	77.9	-0.000	-0.000	1.03	1.02
strong	weak	1	1.00	0.00	100.0	80.4	85.5	84.7	0.000	0.000	1.12	1.02
strong	strong	1	1.00	0.00	99.9	79.2	84.9	84.2	-0.001	-0.001	1.11	1.00

Table 2.8C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.5*SD of logit of propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null: Coverage of 95% CI		Under null: Length FM CI / length AFM CI	Under alternative: Coverage of 95% CI		Under alternative: Length FM CI / length AFM CI
Outcome	Treatment				FM	AFM		FM	AFM	
weak	weak	5	5.00	100.0	0.95	0.95	1.02	0.97	0.97	1.03
weak	strong	5	4.92	99.7	0.95	0.95	1.02	0.97	0.97	1.02
mixed	mixed	5	4.99	100.0	0.97	0.96	1.01	0.98	0.97	1.02
strong	weak	5	5.00	100.0	0.97	0.96	1.12	0.99	0.97	1.10
strong	strong	5	4.92	99.7	0.95	0.95	1.11	0.97	0.97	1.09
weak	weak	2	2.00	100.0	0.96	0.95	1.02	0.97	0.96	1.03
weak	strong	2	2.00	99.8	0.96	0.95	1.02	0.97	0.96	1.02
mixed	mixed	2	2.00	100.0	0.96	0.96	1.01	0.97	0.97	1.02
strong	weak	2	2.00	100.0	0.97	0.95	1.12	0.98	0.97	1.11
strong	strong	2	2.00	99.8	0.97	0.95	1.13	0.98	0.96	1.11
weak	weak	1	1.00	100.0	0.96	0.95	1.02	0.96	0.95	1.01
weak	strong	1	1.00	99.9	0.95	0.95	1.02	0.96	0.96	1.01
mixed	mixed	1	1.00	100.0	0.96	0.96	1.01	0.96	0.96	1.00
strong	weak	1	1.00	100.0	0.97	0.95	1.12	0.97	0.96	1.10
strong	strong	1	1.00	99.9	0.97	0.95	1.13	0.97	0.95	1.11

Table 2.9A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.25 based on raw propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Type I Error (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	Treatment					FM	AFM	AFMCC	FM	AFM	FM	AFM
weak	weak	5	5.00	0.00	100.0	2.2	2.6	2.5	0.000	0.000	1.02	1.00
weak	strong	5	4.97	0.23	99.9	1.6	2.1	2.0	-0.003	-0.002	1.02	1.00
mixed	mixed	5	5.00	0.00	100.0	1.8	1.9	1.9	-0.000	-0.000	1.05	1.04
strong	weak	5	5.00	0.00	100.0	1.3	2.3	2.1	0.000	0.000	1.12	1.00
strong	strong	5	4.97	0.23	99.9	0.6	1.2	1.1	-0.010	-0.007	1.10	0.99
weak	weak	2	2.00	0.00	100.0	2.2	2.5	2.4	-0.000	-0.000	1.01	0.99
weak	strong	2	2.00	0.02	99.9	2.1	2.5	2.3	-0.001	-0.001	1.00	0.98
mixed	mixed	2	2.00	0.00	100.0	1.9	2.0	1.9	0.001	0.001	1.06	1.04
strong	weak	2	2.00	0.00	100.0	1.1	2.2	2.0	-0.000	-0.000	1.14	1.02
strong	strong	2	2.00	0.02	99.9	1.0	2.1	2.0	-0.002	-0.001	1.13	1.01
weak	weak	1	1.00	0.00	100.0	2.4	2.6	2.4	-0.000	-0.000	1.02	0.99
weak	strong	1	1.00	0.00	99.9	2.1	2.3	2.1	-0.001	-0.001	1.02	1.00
mixed	mixed	1	1.00	0.00	100.0	2.0	2.1	2.0	-0.000	-0.000	1.04	1.03
strong	weak	1	1.00	0.00	100.0	1.6	2.7	2.5	-0.000	-0.000	1.12	1.00
strong	strong	1	1.00	0.00	99.9	1.4	2.6	2.4	-0.001	-0.001	1.12	0.99

Table 2.9B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.25 based on raw propensity score, total sample size (pre-match) = 20000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Power (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	5.00	0.00	100.0	91.5	92.2	92.1	-0.000	-0.000	1.13	1.10
weak	strong	5	4.97	0.23	99.9	89.2	90.4	90.2	-0.004	-0.003	1.13	1.10
mixed	mixed	5	5.00	0.00	100.0	92.6	93.1	93.0	0.000	0.000	1.17	1.15
strong	weak	5	5.00	0.00	100.0	93.6	96.1	96.0	0.000	0.000	1.24	1.13
strong	strong	5	4.97	0.23	99.9	86.8	92.9	92.5	-0.010	-0.007	1.20	1.11
weak	weak	2	2.00	0.00	100.0	87.5	88.6	88.2	0.000	0.000	1.10	1.07
weak	strong	2	2.00	0.02	99.9	86.2	87.1	86.6	-0.001	-0.000	1.08	1.05
mixed	mixed	2	2.00	0.00	100.0	87.2	88.0	87.5	0.000	0.000	1.11	1.09
strong	weak	2	2.00	0.00	100.0	88.8	92.6	92.2	-0.001	-0.001	1.20	1.08
strong	strong	2	2.00	0.02	99.9	87.1	91.7	91.1	-0.002	-0.001	1.18	1.06
weak	weak	1	1.00	0.00	100.0	78.0	78.9	78.2	-0.000	-0.000	1.02	1.01
weak	strong	1	1.00	0.00	99.9	78.3	79.1	78.1	-0.000	-0.000	1.03	1.02
mixed	mixed	1	1.00	0.00	100.0	78.8	79.2	78.2	0.000	0.000	1.04	1.03
strong	weak	1	1.00	0.00	100.0	79.8	85.3	84.4	-0.001	-0.001	1.12	1.02
strong	strong	1	1.00	0.00	99.9	79.9	85.3	84.5	-0.001	-0.001	1.13	1.02

Table 2.9C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.25 based on raw propensity score, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null			Under alternative		
					Coverage of 95% CI		Length FM CI / length AFM CI	Coverage of 95% CI		Length FM CI / length AFM CI
Outcome	Treatment				FM	AFM		FM	AFM	
weak	weak	5	5.00	100.0	0.95	0.95	1.02	0.98	0.97	1.03
weak	strong	5	4.97	99.9	0.95	0.95	1.02	0.97	0.97	1.02
mixed	mixed	5	5.00	100.0	0.96	0.96	1.01	0.98	0.98	1.02
strong	weak	5	5.00	100.0	0.97	0.95	1.12	0.99	0.97	1.10
strong	strong	5	4.97	99.9	0.96	0.94	1.12	0.97	0.96	1.10
weak	weak	2	2.00	100.0	0.95	0.95	1.02	0.97	0.96	1.03
weak	strong	2	2.00	99.9	0.95	0.95	1.02	0.96	0.96	1.02
mixed	mixed	2	2.00	100.0	0.96	0.96	1.01	0.97	0.97	1.02
strong	weak	2	2.00	100.0	0.98	0.96	1.12	0.98	0.96	1.11
strong	strong	2	2.00	99.9	0.97	0.95	1.13	0.98	0.96	1.11
weak	weak	1	1.00	100.0	0.95	0.95	1.02	0.95	0.95	1.01
weak	strong	1	1.00	99.9	0.96	0.95	1.02	0.96	0.96	1.01
mixed	mixed	1	1.00	100.0	0.96	0.96	1.01	0.96	0.96	1.00
strong	weak	1	1.00	100.0	0.97	0.95	1.12	0.97	0.95	1.10
strong	strong	1	1.00	99.9	0.97	0.95	1.13	0.98	0.96	1.11

Table 2.10A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, no caliper, total sample size (pre-match) = 20000

Selection Models		*Max # controls per treated*	*Mean # controls per treated*	*St. dev. of # controls per treated*	*Mean % of treated subjects matched*	*Type I Error (%)*			*Bias of Est. RD for treated*		*Mean Est. SE of RD / St. dev. of Empirical Est. RD*	
Outcome	*Treatment*					*FM*	*AFM*	*AFMCC*	*FM*	*AFM*	*FM*	*AFM*
weak	weak	5	5.00	0.00	100.0	2.1	2.5	2.4	-0.000	-0.000	1.01	0.99
weak	strong	5	5.00	0.00	100.0	1.5	1.7	1.6	-0.003	-0.003	1.02	1.01
mixed	mixed	5	5.00	0.00	100.0	1.8	1.9	1.9	-0.000	-0.000	1.05	1.04
strong	weak	5	5.00	0.00	100.0	1.2	2.3	2.2	-0.000	-0.000	1.13	1.01
strong	strong	5	5.00	0.00	100.0	0.6	1.2	1.1	-0.010	-0.010	1.08	0.96
weak	weak	2	2.00	0.00	100.0	2.2	2.5	2.4	-0.000	-0.000	1.02	1.00
weak	strong	2	2.00	0.00	100.0	2.2	2.5	2.4	-0.001	-0.001	1.01	0.99
mixed	mixed	2	2.00	0.00	100.0	2.1	2.3	2.2	0.000	0.000	1.04	1.03
strong	weak	2	2.00	0.00	100.0	1.5	2.5	2.4	0.000	0.000	1.12	1.00
strong	strong	2	2.00	0.00	100.0	1.1	2.2	2.0	-0.002	-0.002	1.13	1.00
weak	weak	1	1.00	0.00	100.0	2.2	2.5	2.3	0.001	0.001	1.02	1.00
weak	strong	1	1.00	0.00	100.0	2.2	2.4	2.2	-0.001	-0.001	1.02	1.00
mixed	mixed	1	1.00	0.00	100.0	2.0	2.1	2.0	0.000	0.000	1.03	1.02
strong	weak	1	1.00	0.00	100.0	1.4	2.6	2.4	-0.000	-0.000	1.13	1.01
strong	strong	1	1.00	0.00	100.0	1.1	2.5	2.3	-0.000	-0.000	1.13	1.00

Table 2.10B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, no caliper, total sample size (pre-match) = 20000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Power (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	5.00	0.00	100.0	92.2	93.0	92.9	0.001	0.001	1.13	1.11
weak	strong	5	5.00	0.00	100.0	89.7	90.5	90.3	-0.003	-0.003	1.12	1.10
mixed	mixed	5	5.00	0.00	100.0	91.8	92.3	92.2	-0.000	-0.000	1.15	1.13
strong	weak	5	5.00	0.00	100.0	93.7	95.9	95.7	0.000	0.000	1.25	1.14
strong	strong	5	5.00	0.00	100.0	87.7	91.6	91.3	-0.010	-0.010	1.20	1.09
weak	weak	2	2.00	0.00	100.0	86.7	87.8	87.4	0.000	0.000	1.07	1.04
weak	strong	2	2.00	0.00	100.0	85.8	87.0	86.5	-0.001	-0.001	1.08	1.06
mixed	mixed	2	2.00	0.00	100.0	87.0	87.6	87.2	-0.000	-0.000	1.11	1.09
strong	weak	2	2.00	0.00	100.0	88.7	92.8	92.2	0.000	0.000	1.19	1.08
strong	strong	2	2.00	0.00	100.0	86.9	91.3	90.8	-0.002	-0.002	1.18	1.06
weak	weak	1	1.00	0.00	100.0	78.0	79.0	78.1	-0.000	-0.000	1.02	1.01
weak	strong	1	1.00	0.00	100.0	77.7	78.6	77.8	-0.001	-0.001	1.03	1.03
mixed	mixed	1	1.00	0.00	100.0	78.2	78.7	77.7	-0.000	-0.000	1.04	1.04
strong	weak	1	1.00	0.00	100.0	80.2	85.6	84.8	0.000	0.000	1.13	1.03
strong	strong	1	1.00	0.00	100.0	80.3	85.9	85.2	-0.001	-0.001	1.13	1.02

Table 2.10C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, no caliper, total sample size (pre-match) = 20000

					Under null			***Under alternative***		
Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***Mean % of treated subjects matched***	***Coverage of 95% CI***		***Length FM CI / length AFM CI***	***Coverage of 95% CI***		***Length FM CI / length AFM CI***
Outcome	***Treatment***				***FM***	***AFM***		***FM***	***AFM***	
weak	weak	5	5.00	100.0	0.95	0.95	1.02	0.97	0.97	1.03
weak	strong	5	5.00	100.0	0.96	0.95	1.02	0.97	0.97	1.02
mixed	mixed	5	5.00	100.0	0.96	0.96	1.01	0.98	0.98	1.02
strong	weak	5	5.00	100.0	0.97	0.95	1.12	0.98	0.97	1.10
strong	strong	5	5.00	100.0	0.96	0.93	1.12	0.97	0.95	1.10
weak	weak	2	2.00	100.0	0.95	0.95	1.02	0.96	0.96	1.03
weak	strong	2	2.00	100.0	0.95	0.95	1.02	0.97	0.96	1.02
mixed	mixed	2	2.00	100.0	0.96	0.96	1.01	0.97	0.97	1.02
strong	weak	2	2.00	100.0	0.97	0.95	1.12	0.98	0.97	1.11
strong	strong	2	2.00	100.0	0.97	0.95	1.13	0.98	0.96	1.12
weak	weak	1	1.00	100.0	0.95	0.95	1.02	0.95	0.95	1.01
weak	strong	1	1.00	100.0	0.96	0.95	1.02	0.96	0.96	1.01
mixed	mixed	1	1.00	100.0	0.96	0.96	1.01	0.96	0.96	1.00
strong	weak	1	1.00	100.0	0.97	0.95	1.12	0.97	0.96	1.10
strong	strong	1	1.00	100.0	0.97	0.95	1.13	0.97	0.95	1.11

Table 2.11A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Type I Error (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	4.56	0.93	99.1	0.9	2.5	2.4	-0.011	-0.000	1.00	0.98
weak	strong	5	3.58	1.59	87.3	0.6	3.6	3.5	-0.018	0.002	1.00	0.95
mixed	mixed	5	4.07	1.35	96.0	0.6	2.4	2.3	-0.013	0.000	1.04	1.02
strong	weak	5	4.56	0.94	99.1	0.1	2.3	2.1	-0.026	-0.001	1.04	1.00
strong	strong	5	3.58	1.58	87.3	0.0	4.2	3.9	-0.042	0.003	1.01	0.93
weak	weak	2	1.99	0.08	99.5	2.1	2.6	2.5	-0.001	-0.000	1.01	0.99
weak	strong	2	1.86	0.34	89.8	1.9	3.3	3.1	-0.004	0.002	1.01	0.99
mixed	mixed	2	1.95	0.21	97.5	1.7	2.2	2.2	-0.002	0.001	1.04	1.03
strong	weak	2	1.99	0.08	99.5	1.2	2.4	2.2	-0.003	-0.000	1.11	1.01
strong	strong	2	1.86	0.34	89.8	0.7	3.2	3.0	-0.013	0.003	1.06	0.98
weak	weak	1	1.00	0.00	99.7	2.3	2.6	2.5	0.000	0.000	1.01	0.99
weak	strong	1	1.00	0.00	93.0	2.5	2.6	2.4	0.002	0.002	1.02	1.01
mixed	mixed	1	1.00	0.00	98.6	2.0	2.1	2.0	0.001	0.001	1.05	1.04
strong	weak	1	1.00	0.00	99.7	1.3	2.2	2.0	-0.000	-0.000	1.13	1.02
strong	strong	1	1.00	0.00	93.1	2.0	2.8	2.5	0.002	0.002	1.08	1.00

Table 2.11B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Power (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	**Treatment**					**FM**	**AFM**	**AFMCC**	**FM**	**AFM**	**FM**	**AFM**
weak	weak	5	4.56	0.93	99.1	83.9	91.5	91.4	-0.011	-0.001	1.12	1.09
weak	strong	5	3.58	1.58	87.3	68.7	86.1	85.7	-0.021	-0.001	1.09	1.01
mixed	mixed	5	4.07	1.34	96.0	81.5	90.9	90.6	-0.013	-0.000	1.15	1.09
strong	weak	5	4.56	0.94	99.1	70.8	95.0	94.7	-0.026	-0.001	1.15	1.10
strong	strong	5	3.58	1.58	87.3	38.0	90.5	90.1	-0.049	-0.004	1.13	1.02
weak	weak	2	1.99	0.08	99.5	86.0	87.4	87.0	-0.001	-0.000	1.08	1.06
weak	strong	2	1.86	0.34	89.9	78.0	83.2	82.5	-0.007	-0.001	1.07	1.03
mixed	mixed	2	1.95	0.21	97.5	84.6	87.0	86.4	-0.003	-0.000	1.10	1.08
strong	weak	2	1.99	0.08	99.5	87.3	92.6	92.2	-0.002	0.001	1.17	1.07
strong	strong	2	1.86	0.34	89.9	73.1	88.8	88.1	-0.018	-0.002	1.14	1.05
weak	weak	1	1.00	0.00	99.7	78.6	79.4	78.6	0.000	0.000	1.03	1.02
weak	strong	1	1.00	0.00	93.0	76.2	76.5	75.5	-0.001	-0.001	1.03	1.03
mixed	mixed	1	1.00	0.00	98.6	77.9	78.1	77.3	-0.001	-0.001	1.04	1.04
strong	weak	1	1.00	0.00	99.7	80.4	85.3	84.4	-0.000	-0.000	1.11	1.02
strong	strong	1	1.00	0.00	93.1	80.5	83.8	82.9	-0.000	-0.000	1.09	1.03

Table 2.11C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null: Coverage of 95% CI		Under null: Length FM CI / length AFM CI	Under alternative: Coverage of 95% CI		Under alternative: Length FM CI / length AFM CI
Outcome	Treatment				FM	AFM		FM	AFM	
weak	weak	5	4.56	99.1	0.94	0.95	1.01	0.96	0.97	1.02
weak	strong	5	3.58	87.3	0.91	0.94	1.00	0.92	0.95	1.01
mixed	mixed	5	4.07	96.0	0.94	0.95	1.00	0.96	0.97	1.01
strong	weak	5	4.56	99.1	0.88	0.95	1.08	0.90	0.97	1.06
strong	strong	5	3.58	87.3	0.73	0.93	1.01	0.65	0.95	1.01
weak	weak	2	1.99	99.5	0.95	0.95	1.02	0.97	0.96	1.03
weak	strong	2	1.86	89.8	0.95	0.95	1.01	0.96	0.96	1.01
mixed	mixed	2	1.95	97.5	0.96	0.96	1.01	0.97	0.97	1.01
strong	weak	2	1.99	99.5	0.97	0.95	1.11	0.98	0.96	1.10
strong	strong	2	1.86	89.8	0.95	0.95	1.05	0.94	0.96	1.05
weak	weak	1	1.00	99.7	0.95	0.95	1.02	0.95	0.95	1.01
weak	strong	1	1.00	93.0	0.95	0.95	1.01	0.96	0.96	1.00
mixed	mixed	1	1.00	98.6	0.96	0.96	1.01	0.96	0.96	1.00
strong	weak	1	1.00	99.7	0.97	0.95	1.11	0.97	0.95	1.09
strong	strong	1	1.00	93.1	0.96	0.95	1.08	0.97	0.96	1.06

Table 2.12A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated normal & bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Type I Error (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	Treatment					FM	AFM	AFMCC	FM	AFM	FM	AFM
weak	weak	5	4.58	0.91	99.2	1.0	2.6	2.5	-0.010	-0.000	1.01	0.99
weak	strong	5	3.60	1.58	87.5	0.6	3.7	3.5	-0.017	0.002	1.00	0.95
mixed	mixed	5	4.11	1.32	96.3	0.6	2.6	2.5	-0.014	0.001	1.03	1.00
strong	weak	5	4.58	0.91	99.2	0.2	2.1	2.0	-0.025	-0.001	1.05	1.00
strong	strong	5	3.60	1.58	87.5	0.0	4.1	3.8	-0.042	0.003	1.03	0.94
weak	weak	2	1.99	0.08	99.5	2.1	2.5	2.5	-0.001	0.000	1.02	1.00
weak	strong	2	1.86	0.34	90.0	1.8	3.0	2.8	-0.004	0.002	1.01	0.99
mixed	mixed	2	1.96	0.20	97.7	1.7	2.3	2.3	-0.002	0.001	1.05	1.03
strong	weak	2	1.99	0.08	99.5	1.1	2.3	2.2	-0.002	-0.000	1.12	1.02
strong	strong	2	1.86	0.34	90.0	0.6	3.3	3.1	-0.012	0.003	1.06	0.99
weak	weak	1	1.00	0.00	99.6	2.4	2.6	2.4	0.000	0.000	1.01	1.00
weak	strong	1	1.00	0.00	93.2	2.6	2.7	2.6	0.002	0.002	1.01	1.00
mixed	mixed	1	1.00	0.00	98.7	2.2	2.3	2.1	0.000	0.000	1.04	1.03
strong	weak	1	1.00	0.00	99.7	1.5	2.4	2.2	-0.000	-0.000	1.12	1.01
strong	strong	1	1.00	0.00	93.1	1.9	2.9	2.6	0.002	0.002	1.08	1.00

Table 2.12B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated normal & bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Power (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	Treatment					FM	AFM	AFMCC	FM	AFM	FM	AFM
weak	weak	5	4.58	0.91	99.2	84.2	91.3	91.1	-0.011	-0.001	1.11	1.08
weak	strong	5	3.60	1.58	87.5	70.1	86.7	86.4	-0.021	-0.001	1.10	1.02
mixed	mixed	5	4.11	1.32	96.3	79.9	91.1	91.0	-0.015	-0.000	1.13	1.08
strong	weak	5	4.58	0.91	99.2	71.6	95.2	94.9	-0.025	-0.001	1.16	1.12
strong	strong	5	3.60	1.58	87.5	39.2	90.4	90.0	-0.049	-0.003	1.13	1.01
weak	weak	2	1.99	0.08	99.5	85.3	87.0	86.5	-0.001	-0.000	1.08	1.05
weak	strong	2	1.86	0.34	90.1	77.9	82.9	82.3	-0.007	-0.001	1.06	1.03
mixed	mixed	2	1.96	0.20	97.8	84.4	87.0	86.6	-0.003	-0.000	1.10	1.08
strong	weak	2	1.99	0.08	99.5	86.5	91.9	91.5	-0.003	-0.000	1.18	1.08
strong	strong	2	1.86	0.34	90.1	73.3	88.7	88.0	-0.018	-0.002	1.14	1.06
weak	weak	1	1.00	0.00	99.7	78.3	79.1	78.4	0.000	0.000	1.03	1.02
weak	strong	1	1.00	0.00	93.2	76.2	76.7	75.9	0.000	0.000	1.00	1.00
mixed	mixed	1	1.00	0.00	98.6	78.2	78.5	77.6	-0.000	-0.000	1.04	1.03
strong	weak	1	1.00	0.00	99.6	80.6	85.6	84.7	0.000	0.000	1.12	1.02
strong	strong	1	1.00	0.00	93.2	79.9	83.5	82.5	-0.001	-0.001	1.08	1.02

Table 2.12C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated normal & bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null: Coverage of 95% CI		Under null: Length FM CI / length AFM CI	Under alternative: Coverage of 95% CI		Under alternative: Length FM CI / length AFM CI
Outcome	*Treatment*				*FM*	*AFM*		*FM*	*AFM*	
weak	weak	5	4.58	99.2	0.94	0.95	1.01	0.96	0.96	1.02
weak	strong	5	3.60	87.5	0.92	0.94	1.00	0.92	0.95	1.01
mixed	mixed	5	4.11	96.3	0.93	0.95	1.00	0.95	0.96	1.01
strong	weak	5	4.58	99.2	0.88	0.95	1.08	0.91	0.97	1.07
strong	strong	5	3.60	87.5	0.73	0.93	1.01	0.65	0.95	1.01
weak	weak	2	1.99	99.5	0.95	0.95	1.02	0.96	0.96	1.02
weak	strong	2	1.86	90.0	0.95	0.95	1.01	0.95	0.95	1.01
mixed	mixed	2	1.96	97.7	0.96	0.96	1.01	0.97	0.96	1.02
strong	weak	2	1.99	99.5	0.97	0.95	1.11	0.98	0.97	1.10
strong	strong	2	1.86	90.0	0.95	0.95	1.05	0.95	0.96	1.05
weak	weak	1	1.00	99.6	0.95	0.95	1.02	0.96	0.96	1.01
weak	strong	1	1.00	93.2	0.96	0.95	1.01	0.95	0.95	1.00
mixed	mixed	1	1.00	98.7	0.96	0.96	1.01	0.96	0.96	1.01
strong	weak	1	1.00	99.7	0.97	0.95	1.11	0.97	0.96	1.09
strong	strong	1	1.00	93.1	0.97	0.95	1.08	0.96	0.95	1.06

Table 2.13A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Type I Error (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	Treatment					FM	AFM	AFMCC	FM	AFM	FM	AFM
weak	weak	5	4.37	1.12	98.3	0.5	2.7	2.6	-0.018	0.000	1.00	0.97
weak	strong	5	3.36	1.65	80.0	0.5	4.4	4.2	-0.020	0.004	0.99	0.94
mixed	mixed	5	3.82	1.49	92.3	0.4	2.6	2.5	-0.016	0.001	1.06	1.02
strong	weak	5	4.37	1.12	98.3	0.0	2.0	1.9	-0.040	-0.001	1.05	1.00
strong	strong	5	3.36	1.65	79.9	0.0	5.6	5.3	-0.039	0.008	1.02	0.94
weak	weak	2	1.98	0.13	99.0	1.8	2.4	2.3	-0.002	0.000	1.03	1.00
weak	strong	2	1.81	0.39	82.5	1.8	3.7	3.5	-0.003	0.005	1.02	0.99
mixed	mixed	2	1.91	0.28	94.5	1.3	2.1	1.9	-0.003	0.001	1.07	1.05
strong	weak	2	1.98	0.13	99.0	0.7	2.1	1.9	-0.006	-0.001	1.16	1.02
strong	strong	2	1.81	0.39	82.5	0.7	4.3	4.0	-0.012	0.007	1.07	0.98
weak	weak	1	1.00	0.00	99.4	2.2	2.5	2.4	0.000	0.000	1.03	1.00
weak	strong	1	1.00	0.00	86.5	3.0	3.1	2.9	0.004	0.004	1.01	1.00
mixed	mixed	1	1.00	0.00	96.6	2.0	2.0	1.9	0.001	0.001	1.06	1.05
strong	weak	1	1.00	0.00	99.4	1.2	2.4	2.3	-0.000	-0.000	1.16	1.01
strong	strong	1	1.00	0.00	86.5	2.4	3.4	3.2	0.006	0.006	1.09	1.00

Table 2.13B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Power (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	4.37	1.12	98.3	77.4	91.7	91.4	-0.018	-0.001	1.12	1.07
weak	strong	5	3.36	1.65	80.0	60.7	83.8	83.3	-0.026	-0.002	1.12	1.02
mixed	mixed	5	3.82	1.49	92.4	74.7	88.9	88.6	-0.019	-0.001	1.14	1.08
strong	weak	5	4.37	1.12	98.3	51.5	95.4	95.1	-0.040	-0.002	1.15	1.09
strong	strong	5	3.36	1.65	79.9	35.3	90.2	89.6	-0.052	-0.004	1.16	1.02
weak	weak	2	1.98	0.13	99.0	85.0	87.9	87.3	-0.003	-0.000	1.08	1.05
weak	strong	2	1.81	0.39	82.5	73.6	80.9	80.2	-0.009	-0.001	1.06	1.02
mixed	mixed	2	1.91	0.28	94.6	82.7	86.5	86.0	-0.004	0.000	1.11	1.09
strong	weak	2	1.98	0.13	99.0	85.6	94.2	93.6	-0.006	-0.000	1.21	1.08
strong	strong	2	1.81	0.39	82.5	69.2	88.4	87.7	-0.023	-0.004	1.15	1.05
weak	weak	1	1.00	0.00	99.4	77.4	79.0	78.1	-0.001	-0.001	1.04	1.02
weak	strong	1	1.00	0.00	86.4	74.3	74.7	73.7	-0.001	-0.001	1.03	1.02
mixed	mixed	1	1.00	0.00	96.6	77.8	78.2	77.3	-0.001	-0.001	1.06	1.05
strong	weak	1	1.00	0.00	99.4	81.6	88.0	87.3	0.000	0.000	1.17	1.03
strong	strong	1	1.00	0.00	86.5	80.6	84.1	83.1	-0.003	-0.003	1.09	1.03

Table 2.13C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		*Max # controls per treated*	*Mean # controls per treated*	*Mean % of treated subjects matched*	*Under null*			*Under alternative*		
					Coverage of 95% CI		*Length FM CI / length AFM CI*	*Coverage of 95% CI*		*Length FM CI / length AFM CI*
Outcome	*Treatment*				*FM*	*AFM*		*FM*	*AFM*	
weak	weak	5	4.37	98.3	0.91	0.94	1.02	0.93	0.96	1.02
weak	strong	5	3.36	80.0	0.91	0.93	0.99	0.90	0.96	1.01
mixed	mixed	5	3.82	92.3	0.93	0.96	1.00	0.94	0.97	1.01
strong	weak	5	4.37	98.3	0.76	0.95	1.09	0.76	0.97	1.07
strong	strong	5	3.36	79.9	0.76	0.93	1.00	0.61	0.95	1.00
weak	weak	2	1.98	99.0	0.96	0.95	1.03	0.97	0.96	1.03
weak	strong	2	1.81	82.5	0.95	0.94	1.01	0.96	0.95	1.01
mixed	mixed	2	1.91	94.5	0.96	0.96	1.01	0.97	0.97	1.02
strong	weak	2	1.98	99.0	0.97	0.96	1.14	0.98	0.96	1.13
strong	strong	2	1.81	82.5	0.95	0.94	1.05	0.92	0.96	1.04
weak	weak	1	1.00	99.4	0.96	0.95	1.03	0.96	0.95	1.02
weak	strong	1	1.00	86.5	0.95	0.95	1.02	0.96	0.96	1.01
mixed	mixed	1	1.00	96.6	0.96	0.96	1.01	0.96	0.96	1.00
strong	weak	1	1.00	99.4	0.98	0.95	1.16	0.98	0.96	1.13
strong	strong	1	1.00	86.5	0.97	0.95	1.08	0.97	0.96	1.06

Table 2.14A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated normal covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Type I Error (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	**Treatment**					**FM**	**AFM**	**AFMCC**	**FM**	**AFM**	**FM**	**AFM**
weak	weak	5	4.05	1.36	95.6	0.2	3.0	2.8	-0.031	0.000	1.01	0.96
weak	strong	5	3.04	1.70	65.1	0.6	5.4	5.1	-0.017	0.009	1.00	0.94
mixed	mixed	5	3.50	1.61	84.9	0.0	3.6	3.4	-0.033	0.004	1.03	0.98
strong	weak	5	4.05	1.36	95.6	0.0	2.1	1.9	-0.063	-0.002	1.08	0.99
strong	strong	5	3.04	1.70	65.1	0.3	11.1	10.3	-0.017	0.020	1.03	0.94
weak	weak	2	1.95	0.22	97.2	1.2	2.7	2.5	-0.006	0.000	1.03	0.98
weak	strong	2	1.73	0.44	67.4	2.2	4.7	4.4	-0.000	0.010	1.01	0.97
mixed	mixed	2	1.84	0.36	87.4	0.8	2.9	2.7	-0.009	0.004	1.08	1.03
strong	weak	2	1.95	0.22	97.2	0.1	1.9	1.7	-0.017	-0.002	1.21	1.02
strong	strong	2	1.73	0.44	67.4	2.2	9.9	9.0	0.004	0.021	1.07	0.98
weak	weak	1	1.00	0.00	98.3	1.9	2.4	2.2	0.000	0.000	1.05	1.00
weak	strong	1	1.00	0.00	71.6	4.0	4.3	4.0	0.010	0.010	1.02	1.00
mixed	mixed	1	1.00	0.00	90.9	1.9	2.3	2.2	0.003	0.003	1.09	1.05
strong	weak	1	1.00	0.00	98.3	0.6	2.4	2.1	-0.001	-0.001	1.27	1.01
strong	strong	1	1.00	0.00	71.6	5.8	7.7	7.0	0.018	0.018	1.08	1.01

Table 2.14B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated normal covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		*Max # controls per treated*	*Mean # controls per treated*	*St. dev. of # controls per treated*	*Mean % of treated subjects matched*	*Power (%)*			*Bias of Est. RD for treated*		*Mean Est. SE of RD / St. dev. of Empirical Est. RD*	
Outcome	*Treatment*					*FM*	*AFM*	*AFMCC*	*FM*	*AFM*	*FM*	*AFM*
weak	weak	5	4.04	1.36	95.6	60.5	91.1	90.8	-0.033	-0.002	1.11	1.03
weak	strong	5	3.04	1.70	65.1	50.9	76.6	75.7	-0.030	-0.004	1.10	1.00
mixed	mixed	5	3.50	1.61	84.9	48.4	89.0	88.6	-0.040	-0.003	1.16	1.05
strong	weak	5	4.05	1.36	95.6	19.9	96.8	96.6	-0.064	-0.003	1.21	1.10
strong	strong	5	3.04	1.70	65.0	48.4	89.5	88.8	-0.044	-0.007	1.23	1.06
weak	weak	2	1.95	0.22	97.2	81.9	88.8	88.3	-0.007	-0.001	1.11	1.06
weak	strong	2	1.73	0.44	67.4	65.1	74.9	73.7	-0.013	-0.003	1.08	1.03
mixed	mixed	2	1.84	0.36	87.4	74.3	86.1	85.5	-0.014	-0.002	1.13	1.07
strong	weak	2	1.95	0.22	97.2	78.5	96.2	95.9	-0.016	-0.002	1.29	1.11
strong	strong	2	1.73	0.44	67.4	72.7	89.1	88.3	-0.023	-0.006	1.22	1.10
weak	weak	1	1.00	0.00	98.3	78.9	81.3	80.4	-0.001	-0.001	1.06	1.02
weak	strong	1	1.00	0.00	71.6	69.8	70.5	69.1	-0.002	-0.002	1.03	1.03
mixed	mixed	1	1.00	0.00	90.9	78.2	80.3	79.4	-0.001	-0.001	1.07	1.05
strong	weak	1	1.00	0.00	98.4	83.8	92.7	92.1	-0.000	-0.000	1.26	1.03
strong	strong	1	1.00	0.00	71.7	84.1	87.0	85.7	-0.005	-0.005	1.15	1.09

Table 2.14C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated normal covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null: Coverage of 95% CI		Under null: Length FM CI / length AFM CI	Under alternative: Coverage of 95% CI		Under alternative: Length FM CI / length AFM CI
Outcome	***Treatment***				***FM***	***AFM***		***FM***	***AFM***	
weak	weak	5	4.05	95.6	0.84	0.94	1.02	0.84	0.96	1.02
weak	strong	5	3.04	65.1	0.93	0.93	0.99	0.89	0.95	1.00
mixed	mixed	5	3.50	84.9	0.83	0.94	1.00	0.79	0.96	1.01
strong	weak	5	4.05	95.6	0.42	0.95	1.11	0.37	0.97	1.09
strong	strong	5	3.04	65.1	0.92	0.88	0.99	0.72	0.95	0.99
weak	weak	2	1.95	97.2	0.95	0.95	1.04	0.97	0.96	1.04
weak	strong	2	1.73	67.4	0.95	0.94	1.01	0.95	0.95	1.01
mixed	mixed	2	1.84	87.4	0.96	0.95	1.03	0.95	0.96	1.03
strong	weak	2	1.95	97.2	0.96	0.96	1.21	0.97	0.97	1.18
strong	strong	2	1.73	67.4	0.96	0.90	1.04	0.93	0.96	1.03
weak	weak	1	1.00	98.3	0.96	0.95	1.05	0.96	0.96	1.04
weak	strong	1	1.00	71.6	0.95	0.94	1.02	0.96	0.96	1.00
mixed	mixed	1	1.00	90.9	0.97	0.96	1.04	0.97	0.96	1.03
strong	weak	1	1.00	98.3	0.99	0.95	1.26	0.99	0.96	1.23
strong	strong	1	1.00	71.6	0.94	0.92	1.08	0.97	0.96	1.05

Table 2.15A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Type I Error (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	4.79	0.59	99.9	1.8	2.5	2.4	-0.003	-0.000	1.00	0.99
weak	strong	5	4.03	1.36	96.5	1.0	3.1	3.0	-0.010	0.001	0.99	0.96
mixed	mixed	5	4.47	1.00	99.4	1.5	2.7	2.6	-0.006	-0.000	1.01	0.98
strong	weak	5	4.79	0.59	99.9	0.9	2.5	2.4	-0.008	-0.000	1.04	0.99
strong	strong	5	4.04	1.36	96.5	0.1	3.2	3.1	-0.029	0.001	1.01	0.96
weak	weak	2	2.00	0.00	100.0	2.3	2.5	2.4	0.000	0.000	1.01	1.00
weak	strong	2	1.95	0.20	98.3	2.0	2.6	2.5	-0.002	0.000	1.01	1.00
mixed	mixed	2	2.00	0.03	99.9	1.9	2.3	2.1	-0.000	-0.000	1.04	1.01
strong	weak	2	2.00	0.00	100.0	1.8	2.6	2.4	-0.000	0.000	1.06	1.00
strong	strong	2	1.95	0.20	98.3	1.3	2.6	2.5	-0.005	0.001	1.05	1.00
weak	weak	1	1.00	0.00	100.0	2.5	2.6	2.4	-0.000	-0.000	1.01	1.00
weak	strong	1	1.00	0.00	99.4	2.5	2.5	2.4	0.000	0.000	1.00	1.00
mixed	mixed	1	1.00	0.00	100.0	2.2	2.5	2.4	-0.000	-0.000	1.02	0.99
strong	weak	1	1.00	0.00	100.0	1.9	2.5	2.3	-0.000	-0.000	1.06	1.00
strong	strong	1	1.00	0.00	99.4	2.1	2.7	2.6	0.000	0.000	1.04	0.99

Table 2.15B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		*Max # controls per treated*	*Mean # controls per treated*	*St. dev. of # controls per treated*	*Mean % of treated subjects matched*	*Power (%)*			*Bias of Est. RD for treated*		*Mean Est. SE of RD / St. dev. of Empirical Est. RD*	
Outcome	*Treatment*					*FM*	*AFM*	*AFMCC*	*FM*	*AFM*	*FM*	*AFM*
weak	weak	5	4.79	0.59	99.9	89.5	91.6	91.4	-0.003	-0.000	1.11	1.09
weak	strong	5	4.03	1.36	96.5	81.8	89.4	89.2	-0.011	-0.000	1.09	1.04
mixed	mixed	5	4.47	0.99	99.4	87.8	92.0	91.8	-0.006	0.000	1.11	1.07
strong	weak	5	4.79	0.59	99.9	87.0	94.1	93.9	-0.008	-0.000	1.14	1.09
strong	strong	5	4.04	1.36	96.5	63.0	91.8	91.4	-0.031	-0.001	1.11	1.04
weak	weak	2	2.00	0.00	100.0	86.1	86.6	86.3	-0.000	-0.000	1.07	1.05
weak	strong	2	1.95	0.20	98.3	84.6	86.1	85.4	-0.002	0.000	1.07	1.05
mixed	mixed	2	2.00	0.03	99.9	86.5	87.9	87.3	-0.000	0.000	1.08	1.04
strong	weak	2	2.00	0.00	100.0	87.0	89.4	88.9	-0.000	-0.000	1.11	1.04
strong	strong	2	1.95	0.20	98.3	83.2	88.8	88.3	-0.005	0.000	1.11	1.06
weak	weak	1	1.00	0.00	100.0	77.8	78.2	77.3	-0.000	-0.000	1.02	1.02
weak	strong	1	1.00	0.00	99.4	77.9	78.3	77.5	0.000	0.000	1.02	1.02
mixed	mixed	1	1.00	0.00	100.0	78.7	79.9	79.1	0.000	0.000	1.03	1.02
strong	weak	1	1.00	0.00	100.0	78.8	81.7	80.9	-0.000	-0.000	1.06	1.01
strong	strong	1	1.00	0.00	99.4	79.0	81.5	80.6	0.001	0.001	1.06	1.02

Table 2.15C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = independent bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null			Under alternative		
					Coverage of 95% CI		Length FM CI / length AFM CI	Coverage of 95% CI		Length FM CI / length AFM CI
Outcome	Treatment				FM	AFM		FM	AFM	
weak	weak	5	4.79	99.9	0.95	0.95	1.01	0.97	0.97	1.02
weak	strong	5	4.03	96.5	0.94	0.94	1.00	0.95	0.96	1.01
mixed	mixed	5	4.47	99.4	0.95	0.95	1.02	0.97	0.96	1.03
strong	weak	5	4.79	99.9	0.95	0.95	1.06	0.97	0.97	1.05
strong	strong	5	4.04	96.5	0.85	0.94	1.02	0.86	0.96	1.02
weak	weak	2	2.00	100.0	0.95	0.95	1.01	0.96	0.96	1.02
weak	strong	2	1.95	98.3	0.95	0.95	1.01	0.96	0.96	1.01
mixed	mixed	2	2.00	99.9	0.96	0.96	1.03	0.97	0.96	1.03
strong	weak	2	2.00	100.0	0.96	0.95	1.07	0.97	0.96	1.06
strong	strong	2	1.95	98.3	0.96	0.95	1.04	0.97	0.96	1.04
weak	weak	1	1.00	100.0	0.95	0.95	1.01	0.95	0.95	1.00
weak	strong	1	1.00	99.4	0.95	0.95	1.01	0.95	0.95	1.00
mixed	mixed	1	1.00	100.0	0.95	0.95	1.03	0.96	0.96	1.02
strong	weak	1	1.00	100.0	0.96	0.95	1.06	0.96	0.95	1.05
strong	strong	1	1.00	99.4	0.96	0.95	1.05	0.96	0.96	1.04

Table 2.16A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Type I Error (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	Treatment					FM	AFM	AFMCC	FM	AFM	FM	AFM
weak	weak	5	4.69	0.69	99.9	1.6	2.5	2.4	-0.004	0.000	1.01	0.99
weak	strong	5	3.63	1.52	94.4	0.7	3.1	3.0	-0.015	0.001	1.00	0.96
mixed	mixed	5	4.25	1.15	99.3	0.8	2.7	2.7	-0.013	-0.001	1.01	0.97
strong	weak	5	4.69	0.69	99.9	0.5	2.4	2.2	-0.012	-0.000	1.04	0.99
strong	strong	5	3.63	1.52	94.4	0.0	3.6	3.5	-0.038	0.002	1.02	0.94
weak	weak	2	2.00	0.00	100.0	2.3	2.5	2.5	-0.000	-0.000	1.01	1.00
weak	strong	2	1.90	0.29	97.3	1.9	2.7	2.6	-0.003	-0.000	1.00	0.99
mixed	mixed	2	2.00	0.01	100.0	2.0	2.6	2.5	0.000	0.000	1.04	1.00
strong	weak	2	2.00	0.00	100.0	1.6	2.6	2.4	0.000	0.000	1.10	1.00
strong	strong	2	1.90	0.30	97.3	1.0	2.8	2.6	-0.008	0.001	1.05	0.99
weak	weak	1	1.00	0.00	100.0	2.3	2.5	2.3	0.000	0.000	1.01	1.00
weak	strong	1	1.00	0.00	99.9	2.3	2.4	2.3	-0.000	-0.000	1.00	1.00
mixed	mixed	1	1.00	0.00	100.0	2.1	2.5	2.3	0.000	0.000	1.04	1.00
strong	weak	1	1.00	0.00	100.0	1.7	2.7	2.6	0.000	0.000	1.08	0.99
strong	strong	1	1.00	0.00	99.9	2.0	2.6	2.4	-0.000	-0.000	1.05	1.00

Table 2.16B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Power (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	4.69	0.69	99.9	88.6	91.7	91.4	-0.005	-0.000	1.11	1.08
weak	strong	5	3.63	1.52	94.4	76.9	88.2	88.0	-0.016	-0.000	1.10	1.03
mixed	mixed	5	4.25	1.15	99.3	83.1	91.8	91.5	-0.012	0.000	1.10	1.05
strong	weak	5	4.69	0.68	99.9	83.4	94.2	93.9	-0.012	-0.000	1.12	1.07
strong	strong	5	3.63	1.52	94.4	48.8	90.2	89.8	-0.040	-0.001	1.10	1.00
weak	weak	2	2.00	0.00	100.0	86.5	87.4	86.9	0.000	0.000	1.07	1.05
weak	strong	2	1.90	0.29	97.3	83.2	85.8	85.3	-0.003	0.000	1.06	1.04
mixed	mixed	2	2.00	0.01	100.0	86.6	88.5	88.0	0.000	0.000	1.09	1.06
strong	weak	2	2.00	0.00	100.0	88.2	91.6	91.1	0.000	0.000	1.15	1.07
strong	strong	2	1.90	0.29	97.3	79.3	88.2	87.5	-0.009	-0.000	1.10	1.04
weak	weak	1	1.00	0.00	100.0	78.8	79.3	78.5	0.001	0.001	1.02	1.01
weak	strong	1	1.00	0.00	99.9	78.1	78.3	77.5	0.000	0.000	1.01	1.02
mixed	mixed	1	1.00	0.00	100.0	78.6	80.1	79.4	-0.000	-0.000	1.03	1.01
strong	weak	1	1.00	0.00	100.0	80.2	84.3	83.4	0.001	0.001	1.08	1.01
strong	strong	1	1.00	0.00	99.9	78.6	80.6	79.7	-0.001	-0.001	1.05	1.01

Table 2.16C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = correlated bernoulli covariates, caliper width of 0.2*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null: Coverage of 95% CI		Under null: Length FM CI / length AFM CI	Under alternative: Coverage of 95% CI		Under alternative: Length FM CI / length AFM CI
Outcome	***Treatment***				***FM***	***AFM***		***FM***	***AFM***	
weak	weak	5	4.69	99.9	0.95	0.95	1.01	0.97	0.97	1.02
weak	strong	5	3.63	94.4	0.93	0.94	1.00	0.94	0.96	1.01
mixed	mixed	5	4.25	99.3	0.93	0.94	1.03	0.95	0.96	1.03
strong	weak	5	4.69	99.9	0.94	0.95	1.08	0.95	0.96	1.07
strong	strong	5	3.63	94.4	0.78	0.94	1.02	0.76	0.95	1.02
weak	weak	2	2.00	100.0	0.95	0.95	1.01	0.96	0.96	1.02
weak	strong	2	1.90	97.3	0.95	0.95	1.01	0.96	0.96	1.01
mixed	mixed	2	2.00	100.0	0.96	0.95	1.04	0.97	0.96	1.04
strong	weak	2	2.00	100.0	0.97	0.95	1.10	0.98	0.96	1.08
strong	strong	2	1.90	97.3	0.96	0.95	1.05	0.96	0.96	1.04
weak	weak	1	1.00	100.0	0.95	0.95	1.01	0.95	0.95	1.00
weak	strong	1	1.00	99.9	0.95	0.95	1.01	0.95	0.95	1.00
mixed	mixed	1	1.00	100.0	0.96	0.95	1.04	0.96	0.95	1.02
strong	weak	1	1.00	100.0	0.97	0.95	1.09	0.97	0.95	1.07
strong	strong	1	1.00	99.9	0.96	0.95	1.06	0.96	0.95	1.04

Table 2.17A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.1 based on raw propensity score, total sample size (pre-match) = 2000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Type I Error (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	4.83	0.57	99.8	0.9	1.4	1.4	-0.012	-0.007	1.01	0.99
weak	strong	5	3.74	1.57	88.4	0.5	3.2	3.1	-0.020	-0.001	1.00	0.95
mixed	mixed	5	4.31	1.22	97.2	0.6	1.6	1.6	-0.014	-0.003	1.04	1.02
strong	weak	5	4.83	0.57	99.8	0.1	0.4	0.4	-0.030	-0.018	1.03	1.00
strong	strong	5	3.74	1.57	88.4	0.0	2.7	2.5	-0.048	-0.003	1.01	0.93
weak	weak	2	2.00	0.03	99.9	2.1	2.5	2.4	-0.000	0.000	1.02	1.00
weak	strong	2	1.87	0.33	90.7	1.9	3.0	2.9	-0.004	0.002	1.00	0.98
mixed	mixed	2	1.96	0.18	98.4	1.8	2.0	2.0	-0.002	-0.000	1.04	1.03
strong	weak	2	2.00	0.02	99.9	1.2	2.3	2.1	-0.002	-0.001	1.12	1.01
strong	strong	2	1.87	0.33	90.7	0.6	2.9	2.7	-0.015	-0.000	1.05	0.98
weak	weak	1	1.00	0.00	99.9	2.2	2.4	2.2	-0.001	-0.001	1.01	0.99
weak	strong	1	1.00	0.00	93.8	2.5	2.7	2.5	0.001	0.001	1.01	1.00
mixed	mixed	1	1.00	0.00	99.2	2.2	2.2	2.1	-0.001	-0.001	1.04	1.02
strong	weak	1	1.00	0.00	99.9	1.1	2.1	1.9	-0.001	-0.001	1.12	1.01
strong	strong	1	1.00	0.00	93.8	1.6	2.2	2.0	0.000	0.000	1.08	1.00

Table 2.17B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.1 based on raw propensity score, total sample size (pre-match) = 2000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Power (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	4.83	0.57	99.8	83.3	88.2	88.0	-0.012	-0.007	1.10	1.08
weak	strong	5	3.74	1.57	88.3	66.9	84.7	84.4	-0.024	-0.005	1.09	1.01
mixed	mixed	5	4.31	1.22	97.2	79.9	88.9	88.6	-0.015	-0.004	1.13	1.09
strong	weak	5	4.83	0.57	99.8	65.5	85.6	85.1	-0.030	-0.018	1.13	1.10
strong	strong	5	3.74	1.57	88.4	33.0	87.2	86.8	-0.054	-0.009	1.13	1.01
weak	weak	2	2.00	0.03	99.9	85.7	86.9	86.6	-0.001	-0.001	1.08	1.05
weak	strong	2	1.87	0.33	90.7	77.7	83.3	82.6	-0.008	-0.002	1.07	1.04
mixed	mixed	2	1.96	0.18	98.4	84.8	86.7	86.2	-0.003	-0.001	1.11	1.09
strong	weak	2	2.00	0.02	99.9	87.3	91.6	91.2	-0.002	-0.001	1.18	1.07
strong	strong	2	1.87	0.33	90.7	71.5	87.2	86.6	-0.019	-0.004	1.14	1.06
weak	weak	1	1.00	0.00	100.0	78.4	79.1	78.3	-0.000	-0.000	1.03	1.02
weak	strong	1	1.00	0.00	93.8	76.1	76.4	75.6	-0.001	-0.001	1.02	1.01
mixed	mixed	1	1.00	0.00	99.2	78.6	79.0	78.2	-0.001	-0.001	1.04	1.04
strong	weak	1	1.00	0.00	99.9	79.7	84.8	84.1	-0.001	-0.001	1.13	1.04
strong	strong	1	1.00	0.00	93.7	79.3	82.9	81.9	-0.002	-0.002	1.09	1.03

Table 2.17C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.1 based on raw propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null: Coverage of 95% CI		Under null: Length FM CI / length AFM CI	Under alternative: Coverage of 95% CI		Under alternative: Length FM CI / length AFM CI
Outcome	***Treatment***				***FM***	***AFM***		***FM***	***AFM***	
weak	weak	5	4.83	99.8	0.93	0.94	1.02	0.95	0.96	1.02
weak	strong	5	3.74	88.4	0.91	0.94	1.00	0.91	0.95	1.01
mixed	mixed	5	4.31	97.2	0.93	0.96	1.00	0.95	0.96	1.01
strong	weak	5	4.83	99.8	0.84	0.90	1.08	0.86	0.92	1.07
strong	strong	5	3.74	88.4	0.66	0.93	1.01	0.58	0.94	1.02
weak	weak	2	2.00	99.9	0.96	0.95	1.02	0.97	0.96	1.03
weak	strong	2	1.87	90.7	0.95	0.95	1.01	0.96	0.96	1.01
mixed	mixed	2	1.96	98.4	0.96	0.96	1.01	0.97	0.97	1.02
strong	weak	2	2.00	99.9	0.97	0.95	1.11	0.98	0.96	1.10
strong	strong	2	1.87	90.7	0.94	0.95	1.05	0.94	0.96	1.05
weak	weak	1	1.00	99.9	0.95	0.95	1.02	0.96	0.95	1.01
weak	strong	1	1.00	93.8	0.95	0.95	1.01	0.95	0.95	1.00
mixed	mixed	1	1.00	99.2	0.96	0.96	1.01	0.96	0.96	1.00
strong	weak	1	1.00	99.9	0.97	0.96	1.11	0.98	0.96	1.09
strong	strong	1	1.00	93.8	0.97	0.95	1.08	0.96	0.95	1.06

Table 2.18A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.5*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Type I Error (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	4.64	0.83	99.5	1.1	2.2	2.2	-0.012	-0.003	1.00	0.98
weak	strong	5	3.70	1.54	90.3	0.4	2.6	2.5	-0.022	-0.004	1.00	0.95
mixed	mixed	5	4.17	1.28	97.2	0.5	1.8	1.8	-0.014	-0.003	1.04	1.02
strong	weak	5	4.64	0.83	99.5	0.2	1.6	1.4	-0.027	-0.006	1.04	1.00
strong	strong	5	3.70	1.54	90.2	0.0	1.3	1.3	-0.055	-0.013	1.01	0.94
weak	weak	2	2.00	0.05	99.8	2.0	2.4	2.3	-0.001	-0.001	1.03	1.01
weak	strong	2	1.89	0.31	92.7	1.5	2.4	2.3	-0.006	-0.001	1.01	0.99
mixed	mixed	2	1.96	0.18	98.6	1.5	1.9	1.8	-0.003	-0.001	1.05	1.04
strong	weak	2	2.00	0.04	99.8	1.1	2.0	1.9	-0.002	-0.001	1.12	1.01
strong	strong	2	1.89	0.31	92.8	0.3	1.4	1.3	-0.021	-0.008	1.06	0.99
weak	weak	1	1.00	0.00	99.9	2.4	2.7	2.6	0.000	0.000	1.02	1.00
weak	strong	1	1.00	0.00	95.5	2.4	2.5	2.3	-0.000	-0.000	1.01	1.00
mixed	mixed	1	1.00	0.00	99.3	2.0	2.1	2.0	0.000	0.000	1.05	1.04
strong	weak	1	1.00	0.00	99.9	1.4	2.4	2.3	-0.000	-0.000	1.13	1.01
strong	strong	1	1.00	0.00	95.5	1.2	1.8	1.6	-0.005	-0.005	1.08	1.00

Table 2.18B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.5*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Power (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	4.64	0.84	99.5	83.6	90.9	90.6	-0.012	-0.003	1.12	1.09
weak	strong	5	3.70	1.54	90.3	65.8	83.3	82.9	-0.025	-0.007	1.08	1.01
mixed	mixed	5	4.18	1.28	97.2	80.6	89.6	89.3	-0.015	-0.003	1.14	1.09
strong	weak	5	4.64	0.83	99.5	69.3	93.3	93.0	-0.027	-0.005	1.15	1.12
strong	strong	5	3.70	1.54	90.3	25.9	80.9	80.3	-0.059	-0.018	1.12	1.01
weak	weak	2	2.00	0.05	99.8	85.7	87.2	86.7	-0.001	-0.000	1.08	1.05
weak	strong	2	1.89	0.31	92.7	76.9	81.6	80.9	-0.009	-0.004	1.06	1.03
mixed	mixed	2	1.96	0.17	98.6	84.5	86.3	86.0	-0.003	-0.001	1.10	1.08
strong	weak	2	2.00	0.05	99.8	87.0	91.5	91.0	-0.003	-0.001	1.18	1.08
strong	strong	2	1.89	0.31	92.7	65.4	82.6	81.9	-0.025	-0.011	1.12	1.04
weak	weak	1	1.00	0.00	99.9	78.2	79.0	78.1	-0.000	-0.000	1.04	1.03
weak	strong	1	1.00	0.00	95.5	76.8	76.9	75.9	-0.001	-0.001	1.01	1.01
mixed	mixed	1	1.00	0.00	99.3	78.5	79.0	78.0	-0.000	-0.000	1.04	1.04
strong	weak	1	1.00	0.00	99.9	80.3	85.3	84.5	-0.001	-0.001	1.12	1.03
strong	strong	1	1.00	0.00	95.5	75.7	79.5	78.5	-0.006	-0.006	1.07	1.01

Table 2.18C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.5*SD of logit of propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null: Coverage of 95% CI		Under null: Length FM CI / length AFM CI	Under alternative: Coverage of 95% CI		Under alternative: Length FM CI / length AFM CI
Outcome	*Treatment*				*FM*	*AFM*		*FM*	*AFM*	
weak	weak	5	4.64	99.5	0.93	0.95	1.01	0.96	0.97	1.02
weak	strong	5	3.70	90.3	0.90	0.94	1.00	0.90	0.95	1.01
mixed	mixed	5	4.17	97.2	0.94	0.95	1.00	0.95	0.97	1.01
strong	weak	5	4.64	99.5	0.87	0.94	1.08	0.89	0.97	1.07
strong	strong	5	3.70	90.2	0.57	0.91	1.01	0.48	0.91	1.02
weak	weak	2	2.00	99.8	0.96	0.95	1.02	0.97	0.96	1.03
weak	strong	2	1.89	92.7	0.95	0.95	1.01	0.96	0.96	1.01
mixed	mixed	2	1.96	98.6	0.96	0.96	1.01	0.97	0.96	1.02
strong	weak	2	2.00	99.8	0.97	0.95	1.11	0.98	0.97	1.10
strong	strong	2	1.89	92.8	0.92	0.94	1.05	0.91	0.95	1.05
weak	weak	1	1.00	99.9	0.96	0.95	1.02	0.96	0.96	1.01
weak	strong	1	1.00	95.5	0.95	0.95	1.01	0.96	0.96	1.00
mixed	mixed	1	1.00	99.3	0.96	0.96	1.01	0.96	0.96	1.00
strong	weak	1	1.00	99.9	0.97	0.95	1.11	0.97	0.95	1.09
strong	strong	1	1.00	95.5	0.97	0.95	1.08	0.96	0.94	1.06

Table 2.19A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.25 based on raw propensity score, total sample size (pre-match) = 2000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Type I Error (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	4.99	0.11	100.0	0.7	0.9	0.9	-0.013	-0.013	1.00	0.98
weak	strong	5	4.16	1.41	93.2	0.2	0.8	0.8	-0.031	-0.016	0.98	0.96
mixed	mixed	5	4.74	0.80	99.2	0.4	0.6	0.6	-0.019	-0.015	1.02	1.02
strong	weak	5	4.99	0.12	100.0	0.1	0.2	0.1	-0.033	-0.032	0.98	0.91
strong	strong	5	4.16	1.41	93.2	0.0	0.1	0.1	-0.072	-0.038	0.99	0.94
weak	weak	2	2.00	0.00	100.0	2.0	2.3	2.2	-0.001	-0.001	1.02	0.99
weak	strong	2	1.92	0.27	94.9	1.2	1.8	1.8	-0.010	-0.006	1.01	0.99
mixed	mixed	2	1.99	0.09	99.7	1.7	1.9	1.7	-0.003	-0.002	1.05	1.04
strong	weak	2	2.00	0.00	100.0	1.0	1.8	1.7	-0.002	-0.002	1.13	1.01
strong	strong	2	1.92	0.27	94.9	0.1	0.6	0.6	-0.030	-0.019	1.04	0.98
weak	weak	1	1.00	0.00	100.0	2.2	2.4	2.2	-0.001	-0.001	1.02	1.00
weak	strong	1	1.00	0.00	97.1	2.0	2.1	1.9	-0.002	-0.002	1.02	1.01
mixed	mixed	1	1.00	0.00	99.9	1.9	2.0	1.8	-0.001	-0.001	1.04	1.03
strong	weak	1	1.00	0.00	100.0	1.3	2.1	2.0	-0.001	-0.001	1.12	1.01
strong	strong	1	1.00	0.00	97.2	0.8	1.3	1.2	-0.010	-0.010	1.09	1.01

Table 2.19B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.25 based on raw propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Power (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	**Treatment**					**FM**	**AFM**	**AFMCC**	**FM**	**AFM**	**FM**	**AFM**
weak	weak	5	4.99	0.12	100.0	82.4	83.9	83.6	-0.014	-0.013	1.09	1.07
weak	strong	5	4.16	1.41	93.2	59.4	75.9	75.6	-0.032	-0.018	1.08	1.02
mixed	mixed	5	4.74	0.80	99.2	76.6	81.7	81.4	-0.020	-0.015	1.14	1.12
strong	weak	5	4.99	0.12	100.0	61.2	69.3	68.5	-0.033	-0.032	1.08	1.02
strong	strong	5	4.17	1.40	93.2	11.3	55.8	54.8	-0.075	-0.042	1.09	1.01
weak	weak	2	2.00	0.00	100.0	86.3	87.4	87.0	-0.001	-0.001	1.07	1.05
weak	strong	2	1.92	0.27	94.9	76.1	80.6	80.1	-0.011	-0.007	1.08	1.05
mixed	mixed	2	1.99	0.09	99.6	84.8	85.9	85.5	-0.003	-0.003	1.10	1.08
strong	weak	2	2.00	0.00	100.0	87.5	91.7	91.3	-0.002	-0.002	1.19	1.08
strong	strong	2	1.92	0.27	94.9	56.3	73.2	72.0	-0.032	-0.022	1.10	1.03
weak	weak	1	1.00	0.00	100.0	78.3	79.1	78.3	0.000	0.000	1.02	1.01
weak	strong	1	1.00	0.00	97.2	75.3	75.6	74.7	-0.003	-0.003	1.01	1.01
mixed	mixed	1	1.00	0.00	99.9	77.5	77.9	77.0	-0.001	-0.001	1.04	1.04
strong	weak	1	1.00	0.00	100.0	79.7	84.6	83.8	-0.001	-0.001	1.12	1.02
strong	strong	1	1.00	0.00	97.2	71.7	75.6	74.7	-0.011	-0.011	1.09	1.02

Table 2.19C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, caliper width of 0.25 based on raw propensity score, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null: Coverage of 95% CI		Under null: Length FM CI / length AFM CI	Under alternative: Coverage of 95% CI		Under alternative: Length FM CI / length AFM CI
Outcome	Treatment				FM	AFM		FM	AFM	
weak	weak	5	4.99	100.0	0.93	0.93	1.02	0.94	0.94	1.02
weak	strong	5	4.16	93.2	0.83	0.91	1.00	0.84	0.91	1.01
mixed	mixed	5	4.74	99.2	0.91	0.93	1.00	0.93	0.94	1.01
strong	weak	5	4.99	100.0	0.80	0.77	1.09	0.82	0.79	1.08
strong	strong	5	4.16	93.2	0.31	0.73	1.01	0.23	0.68	1.02
weak	weak	2	2.00	100.0	0.95	0.95	1.02	0.96	0.96	1.03
weak	strong	2	1.92	94.9	0.94	0.95	1.01	0.95	0.95	1.02
mixed	mixed	2	1.99	99.7	0.96	0.96	1.01	0.97	0.96	1.02
strong	weak	2	2.00	100.0	0.97	0.96	1.11	0.98	0.97	1.10
strong	strong	2	1.92	94.9	0.87	0.90	1.05	0.85	0.89	1.05
weak	weak	1	1.00	100.0	0.95	0.95	1.02	0.96	0.95	1.01
weak	strong	1	1.00	97.1	0.96	0.96	1.01	0.95	0.95	1.00
mixed	mixed	1	1.00	99.9	0.96	0.96	1.01	0.96	0.96	1.00
strong	weak	1	1.00	100.0	0.97	0.95	1.11	0.97	0.96	1.09
strong	strong	1	1.00	97.2	0.96	0.95	1.08	0.95	0.94	1.06

Table 2.20A. Empirical type I error, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = -0.1, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, no caliper, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	St. dev. of # controls per treated	Mean % of treated subjects matched	Type I Error (%)			Bias of Est. RD for treated		Mean Est. SE of RD / St. dev. of Empirical Est. RD	
Outcome	**Treatment**					**FM**	**AFM**	**AFMCC**	**FM**	**AFM**	**FM**	**AFM**
weak	weak	5	5.00	0.00	100.0	0.8	0.9	0.8	-0.014	-0.014	0.99	0.98
weak	strong	5	5.00	0.00	100.0	0.0	0.0	0.0	-0.052	-0.052	0.95	0.95
mixed	mixed	5	5.00	0.00	100.0	0.2	0.2	0.2	-0.023	-0.023	1.01	1.01
strong	weak	5	5.00	0.00	100.0	0.1	0.2	0.1	-0.034	-0.034	0.98	0.90
strong	strong	5	5.00	0.00	100.0	0.0	0.0	0.0	-0.115	-0.115	0.87	0.86
weak	weak	2	2.00	0.00	100.0	2.2	2.4	2.3	-0.001	-0.001	1.02	0.99
weak	strong	2	2.00	0.00	100.0	0.7	0.8	0.7	-0.018	-0.018	1.00	0.99
mixed	mixed	2	2.00	0.00	100.0	1.6	1.7	1.6	-0.004	-0.004	1.04	1.03
strong	weak	2	2.00	0.00	100.0	1.1	2.0	1.9	-0.001	-0.001	1.13	1.02
strong	strong	2	2.00	0.00	100.0	0.0	0.0	0.0	-0.055	-0.055	0.98	0.93
weak	weak	1	1.00	0.00	100.0	2.2	2.4	2.2	0.000	0.000	1.02	1.00
weak	strong	1	1.00	0.00	100.0	1.8	1.9	1.8	-0.007	-0.007	1.00	0.99
mixed	mixed	1	1.00	0.00	100.0	2.1	2.2	2.0	-0.001	-0.001	1.04	1.03
strong	weak	1	1.00	0.00	100.0	1.3	2.2	2.1	-0.001	-0.001	1.13	1.01
strong	strong	1	1.00	0.00	100.0	0.5	0.7	0.7	-0.022	-0.022	1.05	0.97

Table 2.20B. Empirical power, bias of estimated RD for the treated, estimation of empirical standard deviation (STD) of RD where true RD (p_C-p_E) for the treated = 0, NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, no caliper, total sample size (pre-match) = 2000

Selection Models		***Max # controls per treated***	***Mean # controls per treated***	***St. dev. of # controls per treated***	***Mean % of treated subjects matched***	***Power (%)***			***Bias of Est. RD for treated***		***Mean Est. SE of RD / St. dev. of Empirical Est. RD***	
Outcome	***Treatment***					***FM***	***AFM***	***AFMCC***	***FM***	***AFM***	***FM***	***AFM***
weak	weak	5	5.00	0.00	100.0	82.0	83.0	82.7	-0.014	-0.014	1.10	1.07
weak	strong	5	5.00	0.00	100.0	36.7	37.8	37.4	-0.052	-0.052	1.05	1.04
mixed	mixed	5	5.00	0.00	100.0	73.1	73.9	73.5	-0.024	-0.024	1.12	1.10
strong	weak	5	5.00	0.00	100.0	61.1	67.3	66.6	-0.033	-0.033	1.08	1.00
strong	strong	5	5.00	0.00	100.0	0.6	0.8	0.7	-0.115	-0.115	0.96	0.94
weak	weak	2	2.00	0.00	100.0	86.1	87.0	86.6	-0.001	-0.001	1.08	1.05
weak	strong	2	2.00	0.00	100.0	70.4	71.8	71.0	-0.019	-0.019	1.06	1.04
mixed	mixed	2	2.00	0.00	100.0	84.4	85.1	84.6	-0.004	-0.004	1.10	1.08
strong	weak	2	2.00	0.00	100.0	87.1	91.1	90.6	-0.002	-0.002	1.19	1.08
strong	strong	2	2.00	0.00	100.0	29.4	33.7	32.8	-0.055	-0.055	1.03	0.97
weak	weak	1	1.00	0.00	100.0	78.6	79.4	78.5	0.000	0.000	1.03	1.02
weak	strong	1	1.00	0.00	100.0	72.6	73.1	72.0	-0.007	-0.007	1.03	1.02
mixed	mixed	1	1.00	0.00	100.0	77.4	77.7	76.9	-0.001	-0.001	1.04	1.03
strong	weak	1	1.00	0.00	100.0	80.4	85.3	84.5	-0.001	-0.001	1.13	1.04
strong	strong	1	1.00	0.00	100.0	59.7	64.7	63.5	-0.022	-0.022	1.05	0.98

Table 2.20C. Coverage and width of empirical 95% confidence intervals of RD for treated under null (true RD of treated = -0.1) and under alternative (true RD of treated = 0) with NI margin = -0.1, significance level = 2.5%, covariate scenario = independent normal & bernoulli covariates, no caliper, total sample size (pre-match) = 2000

Selection Models		Max # controls per treated	Mean # controls per treated	Mean % of treated subjects matched	Under null: Coverage of 95% CI		Under null: Length FM CI / length AFM CI	Under alternative: Coverage of 95% CI		Under alternative: Length FM CI / length AFM CI
Outcome	Treatment				FM	AFM		FM	AFM	
weak	weak	5	5.00	100.0	0.92	0.92	1.02	0.94	0.94	1.02
weak	strong	5	5.00	100.0	0.59	0.59	1.00	0.59	0.58	1.01
mixed	mixed	5	5.00	100.0	0.89	0.89	1.00	0.91	0.90	1.01
strong	weak	5	5.00	100.0	0.80	0.75	1.09	0.82	0.77	1.08
strong	strong	5	5.00	100.0	0.03	0.02	1.01	0.01	0.01	1.03
weak	weak	2	2.00	100.0	0.95	0.95	1.02	0.96	0.96	1.03
weak	strong	2	2.00	100.0	0.92	0.91	1.01	0.93	0.92	1.02
mixed	mixed	2	2.00	100.0	0.96	0.95	1.01	0.97	0.97	1.02
strong	weak	2	2.00	100.0	0.97	0.95	1.11	0.98	0.96	1.10
strong	strong	2	2.00	100.0	0.61	0.58	1.05	0.59	0.55	1.06
weak	weak	1	1.00	100.0	0.95	0.95	1.02	0.96	0.96	1.01
weak	strong	1	1.00	100.0	0.95	0.94	1.01	0.95	0.95	1.01
mixed	mixed	1	1.00	100.0	0.96	0.96	1.01	0.96	0.96	1.00
strong	weak	1	1.00	100.0	0.97	0.95	1.11	0.97	0.96	1.10
strong	strong	1	1.00	100.0	0.92	0.90	1.08	0.92	0.89	1.07

A.3. Chapter 3 Tables: Assessing Non-Inferiority via Risk Difference in Propensity-Score Matched Studies When There Are More Experimental Subjects than Control Subjects

Table 3.1. Empirical type I error rate and power: K=200, n_k=1, n_k~constant, p_C=0.2, δ_0=-0.1

Correlation Structure	Type I error rate (%)							Power (%)						
r_{CE}	FM	GEE Exc	MOM	O	ICC	AFM	AFM CC	FM	GEE Exc	MOM	O	ICC	AFM	AFM CC
0	2.5	2.6	2.5	2.5	2.5	2.5	2.4	70.3	70.7	70.1	70.0	70.0	69.9	68.8
0.2	1.9*	2.7	2.6	2.6	2.6	2.6	2.4	71.2	75.3	75.1	74.8	74.7	74.9	73.7
0.4	1.3*	2.6	2.6	2.6	2.6	2.6	2.5	72.6	81.0	81.0	80.4	80.1	80.8	79.5
0.6	0.6*	2.4	2.4	2.3	2.1*	2.4	2.1*	74.6	87.9	87.9	87.8	86.9	87.8	86.8
0.8	0.2*	3.1^	3.1^	2.9^	2.3	3.1^	2.6	78.8	96.2	96.2	96.1	95.2	96.1	95.6

Table 3.2. Empirical type I error rate and power: K=200, n_k=2, n_k~constant, p_C=0.2, δ_0=-0.1, more experimental treated subjects than control subjects

Correlation Structure		*Type I error rate (%)*							*Power (%)*						
r_E	r_{CE}	*FM*	*GEE Exc*	*MOM*	*O*	*ICC*	*AFM*	*AFM CC*	*FM*	*GEE Exc*	*MOM*	*O*	*ICC*	*AFM*	*AFM CC*
0	0	2.5	2.2*	2.1*	2.1*	2.1*	2.4	2.3	85.3	83.4	82.9	82.7	83.0	84.3	83.7
0.2	0.2	2.2*	2.6	2.4	2.4	2.5	2.8^	2.6	86.9	87.3	86.8	86.7	86.7	88.2	87.5
0.4	0.4	1.2*	2.2*	2.1*	2.1*	2.1*	2.6	2.3	88.5	91.4	91.1	91.0	90.8	92.0	91.4
0.6	0.6	0.8*	2.7	2.6	2.5	2.4	3.0^	2.7	90.5	95.6	95.5	95.4	95.1	96.1	95.6
0.8	0.8	0.2*	3.0^	2.9^	2.9^	2.4	3.5^	3.0^	94.1	99.1	99.1	99.1	98.8	99.3	99.1
0.4	0.2	2.3	2.4	2.3	2.2*	2.3	2.6	2.4	86.0	85.4	84.9	84.8	84.7	86.1	85.4
0.6	0.2	2.7	2.5	2.4	2.4	2.4	2.7	2.5	85.3	83.4	82.9	82.8	82.8	84.4	83.5
0.6	0.4	1.7*	2.5	2.4	2.4	2.3	2.7	2.5	87.9	89.9	89.5	89.4	89.1	90.5	89.9
0.8	0.2	3.0^	2.3	2.3	2.2*	2.3	2.6	2.3	84.5	81.3	80.7	80.5	80.5	82.0	81.2
0.8	0.4	2.0*	2.5	2.4	2.3	2.3	2.7	2.5	86.9	87.3	87.0	86.9	86.6	88.1	87.3
0.8	0.6	1.0*	2.5	2.4	2.4	2.3	2.9^	2.5	89.4	93.7	93.4	93.4	92.9	94.3	93.7
0.2	0	3.0^	2.3	2.2*	2.1*	2.2*	2.5	2.3	84.5	81.5	81.0	80.8	81.0	82.4	81.7
0.4	0	3.2^	2.3	2.2*	2.2*	2.3	2.5	2.3	84.9	80.7	80.1	80.0	80.1	81.5	80.8
0.8	0	4.1^	2.3	2.2*	2.2*	2.3	2.5	2.4	82.5	75.9	75.3	75.1	75.3	76.7	75.8
0	0.2	1.7*	2.3	2.2*	2.2*	2.3	2.6	2.4	87.2	88.6	88.1	88.1	88.1	89.3	88.7
0	0.4	0.9*	2.2*	2.1*	2.1*	2.1*	2.5	2.3	89.9	94.2	94.0	94.0	93.8	94.7	94.3

Table 3.3. Empirical type I error rate and power: K=200, $n_k \leq 2$, n_k~uniform, p_C=0.2, δ_0=-0.1, more experimental treated subjects than control subjects

Correlation Structure		***Type I error rate (%)***							***Power (%)***						
r_E	r_{CE}	***FM***	***GEE Exc***	***MOM***	***O***	***ICC***	***AFM***	***AFM CC***	***FM***	***GEE Exc***	***MOM***	***O***	***ICC***	***AFM***	***AFM CC***
0	0	2.4	2.2*	2.2*	2.0*	2.2*	2.3	2.2*	80.0	78.9	76.0	75.5	75.7	78.9	78.1
0.2	0.2	2.1*	2.5	2.3	2.2*	2.3	2.5	2.4	81.5	83.1	80.7	80.4	80.3	83.0	82.2
0.4	0.4	1.5*	2.5	2.4	2.4	2.4	2.6	2.3	83.4	88.2	86.4	85.9	85.5	87.9	87.0
0.6	0.6	0.8*	2.6	2.6	2.5	2.4	2.7	2.4	85.0	93.0	91.7	91.8	91.2	92.5	91.7
0.8	0.8	0.3*	2.9^	2.8^	2.8^	2.3	3.0^	2.6	88.7	98.2	97.8	97.7	97.0	97.9	97.5
0.4	0.2	2.1*	2.4	2.3	2.2*	2.2*	2.5	2.3	80.7	81.5	79.8	79.0	78.8	81.3	80.4
0.6	0.2	2.6	2.7	2.5	2.3	2.4	2.7	2.5	80.5	80.3	79.2	77.3	77.2	79.9	78.9
0.6	0.4	1.7*	2.7	2.6	2.5	2.5	2.7	2.5	82.0	85.9	84.5	83.5	83.1	85.4	84.5
0.8	0.2	2.8^	2.4	2.3	2.4	2.4	2.5	2.3	80.0	78.6	78.1	75.5	75.4	78.0	76.9
0.8	0.4	1.8*	2.5	2.4	2.3	2.2*	2.5	2.3	81.8	84.6	83.6	81.8	81.4	83.8	82.8
0.8	0.6	1.1*	2.8^	2.6	2.5	2.4	2.8^	2.5	83.3	91.0	90.1	89.0	88.3	90.2	89.3
0.2	0	2.8^	2.5	2.4	2.2*	2.3	2.5	2.4	79.6	77.7	75.6	74.5	74.6	77.6	76.6
0.4	0	2.9^	2.4	2.3	2.2*	2.3	2.4	2.3	79.1	76.4	74.8	73.1	73.3	76.2	75.3
0.8	0	3.8^	2.6	2.4	2.5	2.5	2.6	2.4	78.2	73.1	72.8	69.5	69.6	72.7	71.6
0	0.2	1.8*	2.4	2.3	2.1*	2.2*	2.5	2.3	81.6	83.9	81.0	81.5	81.4	84.0	83.3
0	0.4	1.1*	2.6	2.4	2.4	2.4	2.7	2.5	83.8	90.0	87.1	88.3	88.0	90.1	89.4

Table 3.4. Empirical type I error rate and power: K=200, n_k≤2, n_k~beta(2,3), p_C=0.2, δ_0=-0.1, more experimental treated subjects than control subjects

Correlation Structure		Type I error rate (%)							Power (%)						
r_E	r_{CE}	FM	GEE Exc	MOM	O	ICC	AFM	AFM CC	FM	GEE Exc	MOM	O	ICC	AFM	AFM CC
0	0	2.6	2.5	2.4	2.2*	2.3	2.5	2.3	77.0	76.3	73.7	72.9	73.0	76.0	75.1
0.2	0.2	1.9*	2.4	2.3	2.3	2.3	2.4	2.2*	78.4	80.7	78.4	77.9	77.8	80.3	79.3
0.4	0.4	1.5*	2.6	2.5	2.5	2.5	2.6	2.4	80.3	86.1	84.1	83.9	83.5	85.6	84.7
0.6	0.6	0.8*	2.9^	2.7	2.7	2.5	2.9^	2.6	81.9	91.7	90.4	90.1	89.2	90.9	90.1
0.8	0.8	0.3*	3.2^	3.0^	2.9^	2.4	3.1^	2.7	85.5	97.5	97.1	96.8	95.9	96.8	96.3
0.4	0.2	2.0*	2.4	2.3	2.1*	2.2*	2.4	2.2*	78.0	79.3	77.9	76.3	76.2	78.9	78.0
0.6	0.2	2.3	2.6	2.4	2.3	2.4	2.5	2.3	77.5	78.3	77.2	75.1	74.9	77.6	76.6
0.6	0.4	1.5*	2.7	2.4	2.4	2.4	2.6	2.4	79.3	84.6	83.3	82.1	81.5	83.7	82.7
0.8	0.2	2.7	2.8^	2.6	2.5	2.5	2.7	2.5	77.3	77.2	76.6	73.7	73.5	76.1	75.0
0.8	0.4	1.8*	2.7	2.6	2.4	2.4	2.6	2.4	78.9	83.2	82.9	80.2	79.7	81.9	80.9
0.8	0.6	1.1*	2.7	2.5	2.6	2.4	2.7	2.4	81.3	90.4	89.6	88.0	87.1	89.1	88.2
0.2	0	2.6	2.4	2.4	2.2*	2.3	2.4	2.2*	76.8	75.5	73.1	72.0	72.2	75.1	74.0
0.4	0	2.8^	2.4	2.3	2.2*	2.3	2.4	2.2*	76.1	73.9	72.3	70.5	70.7	73.4	72.5
0.8	0	3.2^	2.5	2.4	2.2*	2.3	2.4	2.2*	75.5	71.7	71.5	68.1	68.2	70.8	69.6
0	0.2	1.8*	2.5	2.4	2.2*	2.3	2.4	2.2*	79.0	82.0	79.1	79.2	79.1	81.6	80.7
0	0.4	1.2*	2.5	2.5	2.3	2.3	2.6	2.4	80.5	87.6	84.6	85.5	85.1	87.4	86.6

Table 3.5. Empirical type I error rate and power: K=200, n_k=5, n_k~constant, p_C=0.2, δ_0=-0.1, more experimental treated subjects than control subjects

Correlation Structure		***Type I error rate (%)***							***Power (%)***						
r_E	r_{CE}	***FM***	***GEE Exc***	***MOM***	***O***	***ICC***	***AFM***	***AFM CC***	***FM***	***GEE Exc***	***MOM***	***O***	***ICC***	***AFM***	***AFM CC***
0	0	2.8^	2.0*	2.0*	1.9*	2.0*	2.6	2.5	94.2	91.8	91.4	91.3	91.4	93.6	93.3
0.2	0.2	2.0*	2.0*	2.0*	1.9*	2.0*	2.6	2.5	94.8	93.9	93.6	93.5	93.5	95.4	95.1
0.4	0.4	1.4*	2.2*	2.1*	2.0*	2.1*	3.0^	2.7	96.1	96.6	96.4	96.3	96.2	97.6	97.4
0.6	0.6	0.9*	2.6	2.5	2.4	2.3	3.5^	3.1^	97.6	98.7	98.6	98.6	98.5	99.1	99.0
0.8	0.8	0.3*	3.0^	2.9^	2.9^	2.4	4.3^	3.6^	98.9	[98]	99.8	99.8	99.7	99.9	99.8
0.4	0.2	2.8^	2.2*	2.1*	2.1*	2.2*	2.8^	2.6	94.2	92.0	91.7	91.6	91.6	93.7	93.3
0.6	0.2	3.6^	2.5	2.4	2.3	2.4	3.0^	2.8^	92.9	88.4	87.9	87.9	87.8	90.6	90.0
0.6	0.4	2.3	2.4	2.3	2.3	2.3	3.0^	2.8^	95.1	94.2	94.0	94.0	93.8	95.6	95.2
0.8	0.2	4.7^	2.4	2.3	2.2*	2.3	3.0^	2.7	91.5	84.7	84.3	84.2	84.1	87.1	86.4
0.8	0.4	3.3^	2.6	2.5	2.4	2.4	3.2^	3.0^	93.7	90.7	90.4	90.2	90.0	92.5	92.0
0.8	0.6	1.7*	2.6	2.5	2.5	2.4	3.5^	3.0^	95.9	96.3	96.2	96.1	95.9	97.3	97.0
0.2	0	3.5^	2.2*	2.1*	2.1*	2.2*	2.7	2.6	93.6	89.6	89.2	89.1	89.2	91.7	91.4
0.4	0	4.0^	2.1*	2.0*	2.0*	2.1*	2.6	2.5	92.2	86.5	86.1	86.0	86.1	88.8	88.4
0.8	0	6.3^	2.6	2.5	2.4	2.5	3.1^	2.9^	89.9	79.1	78.5	78.4	78.5	81.7	81.0
0	0.2	1.7*	2.3	2.2*	2.1*	2.2*	2.8^	2.7	95.9	96.0	95.8	95.8	95.7	97.1	97.0
0	0.4	0.5*	2.1*	2.0*	2.0*	2.0*	2.9^	2.7	98.2	99.3	99.2	99.2	99.1	99.5	99.5

Table 3.6. Empirical type I error rate and power: K=200, n_k≤5, n_k~uniform, p_C=0.2, δ_0=-0.1, more experimental treated subjects than control subjects

Correlation Structure		***Type I error rate (%)***							***Power (%)***						
r_E	r_{CE}	***FM***	***GEE Exc***	***MOM***	***O***	***ICC***	***AFM***	***AFM CC***	***FM***	***GEE Exc***	***MOM***	***O***	***ICC***	***AFM***	***AFM CC***
0	0	2.5	2.0*	2.1*	1.8*	1.9*	2.4	2.2*	90.0	87.6	83.8	82.1	82.4	88.8	88.3
0.2	0.2	2.4	2.5	2.3	2.0*	2.1*	2.8^	2.6	91.4	90.9	87.9	86.6	86.6	91.8	91.3
0.4	0.4	1.6*	2.3	2.3	2.0*	2.1*	2.7	2.5	92.3	93.8	91.8	90.6	90.4	94.1	93.7
0.6	0.6	1.1*	2.5	2.4	2.3	2.3	2.9^	2.5	94.0	97.2	96.0	95.2	94.7	96.9	96.5
0.8	0.8	0.6*	3.0^	2.9^	2.9^	2.5	3.3^	2.9^	95.8	99.5	99.2	99.0	98.7	99.1	99.0
0.4	0.2	2.8^	2.4	2.3	2.2*	2.2*	2.7	2.5	90.0	88.1	85.6	83.1	83.0	88.7	88.1
0.6	0.2	3.7^	2.6	2.4	2.3	2.4	2.9^	2.7	89.3	85.4	83.9	80.1	80.1	85.7	84.9
0.6	0.4	2.4	2.5	2.3	2.2*	2.2*	2.8^	2.6	91.2	91.4	89.8	87.2	86.8	91.3	90.7
0.8	0.2	4.5^	2.5	2.3	2.3	2.4	2.8^	2.6	87.8	82.0	81.1	75.4	75.3	81.5	80.6
0.8	0.4	3.3^	2.5	2.3	2.4	2.4	2.8^	2.5	89.6	88.1	87.5	82.7	82.3	87.1	86.3
0.8	0.6	1.9*	2.6	2.6	2.5	2.4	2.8^	2.6	92.0	94.6	93.8	91.3	90.6	93.6	92.9
0.2	0	3.3^	2.2*	2.1*	2.0*	2.1*	2.5	2.3	89.6	85.7	82.3	79.5	79.7	86.9	86.2
0.4	0	4.0^	2.4	2.3	2.2*	2.3	2.7	2.6	88.8	83.0	81.1	76.9	77.1	84.0	83.3
0.8	0	6.0^	2.6	2.4	2.3	2.4	2.9^	2.7	85.9	76.2	75.6	69.0	69.2	76.0	75.1
0	0.2	1.6*	2.2*	2.2*	1.8*	1.9*	2.5	2.4	92.4	93.0	89.4	89.1	89.0	93.9	93.6
0	0.4	0.8*	2.3	2.4	2.1*	2.1*	2.7	2.5	95.1	97.7	94.9	95.8	95.6	98.0	97.9

Table 3.7. Empirical type I error rate and power: K=200, n_k≤5, n_k~beta(2,3), p_C=0.2, δ_0=-0.1, more experimental treated subjects than control subjects

Correlation Structure		***Type I error rate (%)***							***Power (%)***						
r_E	r_{CE}	***FM***	***GEE Exc***	***MOM***	***O***	***ICC***	***AFM***	***AFM CC***	***FM***	***GEE Exc***	***MOM***	***O***	***ICC***	***AFM***	***AFM CC***
0	0	2.6	2.2*	2.2*	1.8*	2.0*	2.4	2.3	88.9	86.8	83.1	82.1	82.3	87.7	87.2
0.2	0.2	2.1*	2.3	2.3	2.0*	2.1*	2.6	2.4	89.8	89.6	87.3	86.1	86.0	90.4	89.8
0.4	0.4	1.7*	2.6	2.3	2.3	2.3	2.8^	2.6	90.5	92.8	90.8	90.0	89.7	93.0	92.4
0.6	0.6	0.9*	2.6	2.5	2.4	2.3	3.0^	2.7	92.5	96.5	95.4	94.9	94.4	96.4	96.0
0.8	0.8	0.6*	3.0^	3.0^	2.8^	2.4	3.2^	2.8^	94.9	99.3	99.0	98.9	98.5	99.0	98.8
0.4	0.2	2.5	2.3	2.2*	2.0*	2.1*	2.5	2.3	88.2	86.8	84.8	82.7	82.6	87.3	86.6
0.6	0.2	3.3^	2.4	2.3	2.1*	2.2*	2.7	2.4	87.6	84.7	83.3	79.8	79.8	84.8	84.0
0.6	0.4	2.3	2.5	2.3	2.3	2.3	2.8^	2.5	89.5	90.5	89.1	87.1	86.7	90.5	89.7
0.8	0.2	4.1^	2.4	2.4	2.2*	2.3	2.6	2.4	86.2	81.4	80.5	76.1	76.0	80.8	79.9
0.8	0.4	3.1^	2.7	2.5	2.5	2.5	3.0^	2.8^	88.0	87.3	86.7	83.1	82.6	86.6	85.7
0.8	0.6	1.6*	2.4	2.5	2.3	2.2*	2.8^	2.5	90.6	94.1	93.5	91.6	90.8	93.3	92.7
0.2	0	3.1^	2.3	2.2*	2.1*	2.2*	2.6	2.4	87.2	83.6	81.0	78.9	79.0	84.6	83.9
0.4	0	3.5^	2.4	2.2*	2.0*	2.1*	2.6	2.4	86.8	81.6	79.9	76.8	77.0	82.3	81.5
0.8	0	5.5^	2.7	2.5	2.5	2.6	3.0^	2.9^	85.1	76.3	75.8	70.6	70.8	76.2	75.2
0	0.2	1.7*	2.3	2.3	2.1*	2.2*	2.6	2.4	90.1	91.2	88.2	87.9	87.8	92.1	91.6
0	0.4	0.8*	2.3	2.3	2.1*	2.1*	2.7	2.4	93.1	96.5	94.0	94.7	94.5	97.0	96.7

Table 3.8A. Empirical type I error, bias, and estimation of SE of RD when true RD for treated (p_C-p_E) = -0.05, event probability under control = 0.2, NI margin = -0.05, signif. level = 2.5%, covariate scenario = indep. normal & bernoulli, pre-match sample allocation (control:treated) = 1:1.5, pre-match total sample size = 2500 where 60% are treated with experimental, caliper width of 0.2*SD of logit of propensity score, matching each control to one or more treated subjects without replacement

Selection Models		Max # treated per control	Mean # treated per control	STD of # treated per control	Mean % of treated subjects matched	Type I Error (%)		Abs. Bias of Est. RD for treated		Mean Est. SE of RD / STD of Empirical RD	
Outcome	Treatment					FM	ICC	FM	ICC	FM	ICC
weak	weak	5	1.97	1.06	99.1	0.1	2.3	-0.021	-0.000	1.03	1.01
weak	strong	5	2.22	1.42	83.4	0.1	2.3	-0.022	0.001	1.03	1.01
mixed	mixed	5	2.12	1.27	94.9	0.1	1.7	-0.019	-0.000	1.07	1.05
strong	weak	5	1.97	1.06	99.1	0.0	1.8	-0.046	-0.002	1.09	1.01
strong	strong	5	2.22	1.42	83.4	0.0	3.3	-0.042	0.003	1.07	0.99
weak	weak	2	1.61	0.49	89.2	0.6	3.1	-0.009	0.002	1.02	1.00
weak	strong	2	1.56	0.50	62.9	1.4	4.0	-0.005	0.005	1.01	1.00
mixed	mixed	2	1.58	0.49	77.0	0.7	2.3	-0.007	0.002	1.06	1.05
strong	weak	2	1.61	0.49	89.2	0.0	3.4	-0.018	0.003	1.10	1.01
strong	strong	2	1.56	0.50	62.9	1.4	10.5	-0.003	0.013	1.04	1.00
weak	weak	1	1.00	0.00	61.4	4.1	4.5	0.005	0.005	1.02	1.00
weak	strong	1	1.00	0.00	44.1	5.6	5.7	0.008	0.008	1.00	1.00
mixed	mixed	1	1.00	0.00	53.7	3.4	3.6	0.004	0.004	1.04	1.04
strong	weak	1	1.00	0.00	61.4	6.5	9.5	0.010	0.010	1.11	1.02
strong	strong	1	1.00	0.00	44.1	18.8	21.3	0.019	0.019	1.06	1.01

Table 3.8B. Empirical power, bias, and estimation of SE of RD when true RD for treated (p_C-p_E) = 0, event probability under control = 0.2, NI margin = -0.05, signif. level = 2.5%, covariate scenario = indep. normal & bernoulli, pre-match sample allocation (control:treated) = 1:1.5, pre-match total sample size = 2500 where 60% are treated with experimental, caliper width of 0.2*SD of logit of propensity score, matching each control to one or more treated subjects without replacement

Selection Models		*Max # treated per control*	*Mean # treated per control*	*STD of # treated per control*	*Mean % of treated subjects matched*	*Power (%)*		*Abs. Bias of Est. RD for treated*		*Mean Est. SE of RD / STD of Empirical RD*	
Outcome	*Treatment*					*FM*	*ICC*	*FM*	*ICC*	*FM*	*ICC*
weak	weak	5	1.97	1.06	99.1	38.2	71.1	-0.022	-0.001	1.00	1.01
weak	strong	5	2.22	1.42	83.4	27.7	57.9	-0.025	-0.001	0.98	1.00
mixed	mixed	5	2.12	1.27	94.9	39.6	66.2	-0.020	-0.000	1.01	1.04
strong	weak	5	1.97	1.06	99.1	3.5	75.1	-0.046	-0.002	1.05	1.02
strong	strong	5	2.22	1.42	83.5	2.9	62.2	-0.049	-0.003	1.02	0.99
weak	weak	2	1.61	0.49	89.2	63.4	80.8	-0.011	-0.000	1.00	1.00
weak	strong	2	1.56	0.50	62.9	57.5	69.9	-0.010	-0.001	1.00	1.01
mixed	mixed	2	1.58	0.49	77.0	64.3	75.7	-0.009	-0.000	1.02	1.03
strong	weak	2	1.61	0.49	89.2	44.4	87.1	-0.021	-0.001	1.07	1.01
strong	strong	2	1.56	0.50	62.9	58.5	84.2	-0.017	-0.002	1.03	1.02
weak	weak	1	1.00	0.00	61.5	82.2	83.0	-0.000	-0.000	1.02	1.00
weak	strong	1	1.00	0.00	44.1	71.2	71.6	-0.000	-0.000	1.02	1.01
mixed	mixed	1	1.00	0.00	53.6	76.4	76.7	-0.000	-0.000	1.04	1.04
strong	weak	1	1.00	0.00	61.5	89.5	92.3	-0.000	-0.000	1.10	1.01
strong	strong	1	1.00	0.00	44.1	88.7	90.2	-0.000	-0.000	1.07	1.03

Table 3.9A. Empirical type I error, bias, and estimation of SE of RD when true RD for treated (p_C-p_E) = -0.05, event probability under control = 0.2, NI margin = -0.05, signif. level = 2.5%, covariate scenario = indep. normal & bernoulli, pre-match sample allocation (control:treated) = 1:2, pre-match total sample size = 3000 where 66.7% are treated with experimental, caliper width of 0.2*SD of logit of propensity score, matching each control to one or more treated subjects without replacement

Selection Models		Max # treated per control	Mean # treated per control	STD of # treated per control	Mean % of treated subjects matched	Type I Error (%)		Abs. Bias of Est. RD for treated		Mean Est. SE of RD / STD of Empirical RD	
Outcome	*Treatment*					*FM*	*ICC*	*FM*	*ICC*	*FM*	*ICC*
weak	weak	5	2.34	1.24	98.3	0.0	1.9	-0.024	-0.000	1.04	1.02
weak	strong	5	2.45	1.51	77.5	0.1	2.8	-0.020	0.003	1.03	1.01
mixed	mixed	5	2.43	1.41	91.1	0.1	2.0	-0.019	0.001	1.05	1.04
strong	weak	5	2.34	1.24	98.3	0.0	1.9	-0.049	-0.001	1.09	1.00
strong	strong	5	2.45	1.51	77.5	0.0	4.5	-0.035	0.007	1.08	1.01
weak	weak	2	1.75	0.43	79.5	1.1	3.5	-0.005	0.003	1.02	1.00
weak	strong	2	1.62	0.48	54.9	1.8	4.8	-0.002	0.006	1.01	1.00
mixed	mixed	2	1.68	0.47	67.7	1.1	3.1	-0.004	0.003	1.05	1.04
strong	weak	2	1.75	0.43	79.5	0.3	5.4	-0.008	0.006	1.11	1.03
strong	strong	2	1.62	0.48	54.8	3.8	16.7	0.003	0.016	1.05	1.02
weak	weak	1	1.00	0.00	48.4	4.9	5.3	0.006	0.006	1.02	1.00
weak	strong	1	1.00	0.00	36.6	6.6	6.9	0.009	0.009	1.02	1.01
mixed	mixed	1	1.00	0.00	43.7	4.2	4.2	0.005	0.005	1.03	1.03
strong	weak	1	1.00	0.00	48.4	8.5	11.9	0.012	0.012	1.11	1.02
strong	strong	1	1.00	0.00	36.6	28.6	31.8	0.022	0.022	1.06	1.01

Table 3.9B. Empirical power, bias, and estimation of SE of RD when true RD for treated (p_C-p_E) = 0, event probability under control = 0.2, NI margin = -0.05, signif. level = 2.5%, covariate scenario = indep. normal & bernoulli, pre-match sample allocation (control:treated) = 1:2, pre-match total sample size = 3000 where 66.7% are treated with experimental, caliper width of 0.2*SD of logit of propensity score, matching each control to one or more treated subjects without replacement

Selection Models		Max # treated per control	Mean # treated per control	STD of # treated per control	Mean % of treated subjects matched	Power (%)		Abs. Bias of Est. RD for treated		Mean Est. SE of RD / STD of Empirical RD	
Outcome	***Treatment***					***FM***	***ICC***	***FM***	***ICC***	***FM***	***ICC***
weak	weak	5	2.34	1.24	98.3	38.7	78.6	-0.024	-0.001	1.00	1.01
weak	strong	5	2.45	1.51	77.5	34.0	66.1	-0.025	-0.001	0.98	1.00
mixed	mixed	5	2.43	1.41	91.2	45.1	74.6	-0.020	-0.001	1.02	1.05
strong	weak	5	2.34	1.24	98.3	2.1	81.1	-0.049	-0.002	1.06	1.01
strong	strong	5	2.45	1.51	77.5	6.5	74.8	-0.043	-0.002	1.03	1.02
weak	weak	2	1.75	0.43	79.5	75.8	87.0	-0.009	-0.001	0.99	1.01
weak	strong	2	1.62	0.48	54.9	65.5	76.8	-0.009	-0.001	0.97	0.99
mixed	mixed	2	1.68	0.47	67.7	73.8	82.6	-0.008	-0.000	1.02	1.04
strong	weak	2	1.75	0.43	79.5	71.9	94.0	-0.015	-0.000	1.07	1.02
strong	strong	2	1.62	0.48	54.9	76.6	92.5	-0.014	-0.001	1.03	1.03
weak	weak	1	1.00	0.00	48.4	84.0	84.6	-0.000	-0.000	1.03	1.01
weak	strong	1	1.00	0.00	36.6	76.6	77.2	-0.000	-0.000	1.01	1.01
mixed	mixed	1	1.00	0.00	43.7	80.4	80.7	-0.000	-0.000	1.04	1.04
strong	weak	1	1.00	0.00	48.4	92.3	94.4	0.000	0.000	1.10	1.01
strong	strong	1	1.00	0.00	36.6	93.8	94.8	-0.000	-0.000	1.08	1.04

Table 3.10A. Empirical type I error, bias, and estimation of SE of RD when true RD for treated (p_C-p_E) = -0.05, event probability under control = 0.2, NI margin = -0.05, signif. level = 2.5%, covariate scenario = indep. normal & bernoulli, pre-match sample allocation (control:treated) = 1:4, pre-match total sample size = 5000 where 80% are treated with experimental, caliper width of 0.2*SD of logit of propensity score, matching each control to one or more treated subjects without replacement

Selection Models		***Max # treated per control***	***Mean # treated per control***	***STD of # treated per control***	***Mean % of treated subjects matched***	***Type I Error (%)***		***Abs. Bias of Est. RD for treated***		***Mean Est. SE of RD / STD of Empirical RD***	
Outcome	***Treatment***					***FM***	***ICC***	***FM***	***ICC***	***FM***	***ICC***
weak	weak	5	3.62	1.39	87.3	0.1	2.7	-0.017	0.002	1.04	1.00
weak	strong	5	3.06	1.62	59.6	0.3	5.0	-0.013	0.006	1.02	0.99
mixed	mixed	5	3.32	1.54	74.4	0.2	2.9	-0.013	0.003	1.05	1.03
strong	weak	5	3.62	1.39	87.3	0.0	4.1	-0.028	0.004	1.12	1.01
strong	strong	5	3.06	1.63	59.7	0.0	16.4	-0.013	0.015	1.08	1.02
weak	weak	2	1.96	0.19	48.4	4.6	6.1	0.004	0.006	1.02	1.00
weak	strong	2	1.76	0.42	36.2	4.5	8.3	0.004	0.010	1.00	1.00
mixed	mixed	2	1.86	0.34	43.6	3.0	4.7	0.002	0.006	1.05	1.04
strong	weak	2	1.96	0.19	48.4	9.0	14.8	0.009	0.012	1.09	1.01
strong	strong	2	1.76	0.42	36.2	25.9	46.0	0.016	0.023	1.04	1.02
weak	weak	1	1.00	0.00	24.9	5.4	5.8	0.007	0.007	1.03	1.02
weak	strong	1	1.00	0.00	21.7	9.9	10.2	0.012	0.012	1.02	1.01
mixed	mixed	1	1.00	0.00	24.2	5.1	5.2	0.007	0.007	1.05	1.04
strong	weak	1	1.00	0.00	24.9	10.6	14.4	0.013	0.013	1.10	1.01
strong	strong	1	1.00	0.00	21.7	52.3	56.5	0.026	0.026	1.07	1.01

Table 3.10B. Empirical power, bias, and estimation of SE of RD when true RD for treated (p_C-p_E) = 0, event probability under control = 0.2, NI margin = -0.05, signif. level = 2.5%, covariate scenario = indep. normal & bernoulli, pre-match sample allocation (control:treated) = 1:4, pre-match total sample size = 5000 where 80% are treated with experimental, caliper width of 0.2*SD of logit of propensity score, matching each control to one or more treated subjects without replacement

Selection Models		***Max # treated per control***	***Mean # treated per control***	***STD of # treated per control***	***Mean % of treated subjects matched***	***Power (%)***		***Abs. Bias of Est. RD for treated***		***Mean Est. SE of RD / STD of Empirical RD***	
Outcome	***Treatment***					***FM***	***ICC***	***FM***	***ICC***	***FM***	***ICC***
weak	weak	5	3.63	1.39	87.3	65.2	92.3	-0.019	-0.000	0.97	1.00
weak	strong	5	3.06	1.62	59.7	57.4	84.3	-0.020	-0.001	0.96	1.00
mixed	mixed	5	3.32	1.54	74.4	68.4	88.6	-0.017	-0.001	0.99	1.04
strong	weak	5	3.62	1.39	87.3	28.8	96.1	-0.033	-0.001	1.05	1.02
strong	strong	5	3.06	1.62	59.7	50.9	95.4	-0.029	-0.001	0.98	1.02
weak	weak	2	1.96	0.19	48.5	92.5	93.8	-0.002	-0.000	0.98	1.00
weak	strong	2	1.76	0.42	36.2	83.9	89.6	-0.007	-0.001	0.98	1.01
mixed	mixed	2	1.86	0.34	43.6	89.0	92.1	-0.004	-0.000	1.01	1.04
strong	weak	2	1.96	0.19	48.4	97.8	98.9	-0.003	-0.000	1.07	1.03
strong	strong	2	1.76	0.42	36.2	97.7	99.2	-0.008	-0.001	1.01	1.04
weak	weak	1	1.00	0.00	24.9	86.0	86.7	-0.000	-0.000	1.02	1.00
weak	strong	1	1.00	0.00	21.7	85.7	86.0	-0.000	-0.000	1.02	1.01
mixed	mixed	1	1.00	0.00	24.2	85.5	85.8	-0.000	-0.000	1.04	1.03
strong	weak	1	1.00	0.00	24.9	93.5	95.3	-0.000	-0.000	1.11	1.02
strong	strong	1	1.00	0.00	21.7	98.9	99.1	-0.000	-0.000	1.11	1.06

Table 3.11A. Empirical type I error, bias, and estimation of SE of RD when true RD for treated (p_C-p_E) = -0.05, event probability under control = 0.2, NI margin = -0.05, signif. level = 2.5%, covariate scenario = indep. normal & bernoulli, caliper width of 0.2*SD of logit of propensity score, Pre-Match number of controls = 1000, matching each treated subject to a control with replacement

Selection Models		Pre Match SS Ratio (treated / cont)	Pre Match Total Sample Size	Max # treated per control	Mean # treated per control	Std. dev. of # treated per control	Mean % of treated subjects matched	Type I Error (%)		Abs. Bias of Est. RD for treated		Mean Est. SE of RD / Std. dev. of Empirical RD	
Outcome	Treatment							FM	ICC	FM	ICC	FM	ICC
weak	weak	1.5	2500	15	2.38	1.90	99.7	0.1	2.1	-0.022	0.000	1.03	1.00
weak	strong	1.5	2500	43	3.07	4.16	97.8	0.0	1.6	-0.039	0.000	1.05	1.01
mixed	mixed	1.5	2500	24	2.63	2.62	99.2	0.1	1.8	-0.024	0.000	1.04	1.02
strong	weak	1.5	2500	15	2.38	1.90	99.7	0.0	2.3	-0.046	0.000	1.09	1.02
strong	strong	1.5	2500	43	3.07	4.15	97.8	0.0	2.4	-0.088	-0.000	1.11	1.00
weak	weak	2	3000	20	2.84	2.40	99.7	0.0	1.9	-0.025	-0.000	1.04	1.01
weak	strong	2	3000	57	3.63	5.32	97.7	0.0	1.6	-0.043	0.001	1.05	1.00
mixed	mixed	2	3000	32	3.12	3.33	99.2	0.0	1.8	-0.026	0.000	1.04	1.01
strong	weak	2	3000	20	2.84	2.39	99.7	0.0	2.2	-0.051	0.000	1.08	1.01
strong	strong	2	3000	57	3.63	5.32	97.7	0.0	2.5	-0.095	-0.001	1.14	1.01
weak	weak	4	5000	37	4.70	4.36	99.6	0.0	2.0	-0.030	0.000	1.05	1.00
weak	strong	4	5000	119	5.76	9.99	97.5	0.0	1.4	-0.053	0.000	1.08	1.00
mixed	mixed	4	5000	62	5.07	6.15	99.1	0.0	1.8	-0.033	0.000	1.06	1.01
strong	weak	4	5000	37	4.70	4.36	99.7	0.0	2.3	-0.061	-0.000	1.12	1.01
strong	strong	4	5000	119	5.77	10.00	97.5	0.0	2.7	-0.108	0.000	1.15	1.00

Table 3.11B. Empirical power, bias, and estimation of SE of RD when true RD for treated (p_C-p_E) = 0, event probability under control = 0.2, NI margin = -0.05, signif. level = 2.5%, covariate scenario = indep. normal & bernoulli, caliper width of 0.2*SD of logit of propensity score, Pre-Match number of controls = 1000, matching each treated subject to a control with replacement

Selection Models		Pre Match SS Ratio (treated / cont)	Pre Match Total Sample Size	Max # treated per control	Mean # treated per control	Std. dev. of # treated per control	Mean % of treated subjects matched	Power (%)		Abs. Bias of Est. RD for treated		Mean Est. SE of RD / Std. dev. of Empirical RD	
Outcome	Treatment							FM	ICC	FM	ICC	FM	ICC
weak	weak	1.5	2500	15	2.38	1.90	99.7	34.0	57.0	-0.022	-0.000	0.99	1.01
weak	strong	1.5	2500	43	3.07	4.15	97.8	7.9	27.2	-0.040	-0.000	0.98	0.99
mixed	mixed	1.5	2500	24	2.63	2.61	99.2	28.9	45.2	-0.024	-0.000	1.00	1.02
strong	weak	1.5	2500	15	2.38	1.90	99.7	3.7	62.3	-0.046	0.000	1.04	1.01
strong	strong	1.5	2500	43	3.07	4.15	97.8	0.0	29.2	-0.088	-0.000	1.05	1.00
weak	weak	2	3000	20	2.84	2.40	99.7	33.7	61.6	-0.025	-0.000	0.98	1.00
weak	strong	2	3000	57	3.63	5.32	97.7	5.5	27.4	-0.044	-0.000	1.00	1.00
mixed	mixed	2	3000	32	3.12	3.33	99.2	27.8	47.7	-0.027	-0.000	0.99	1.01
strong	weak	2	3000	20	2.84	2.40	99.7	2.0	65.6	-0.051	-0.000	1.04	1.01
strong	strong	2	3000	57	3.63	5.32	97.7	0.0	29.4	-0.095	-0.001	1.06	1.00
weak	weak	4	5000	37	4.70	4.36	99.6	29.3	70.4	-0.030	0.000	0.98	1.01
weak	strong	4	5000	118	5.77	9.99	97.5	1.7	27.4	-0.053	-0.000	1.00	1.00
mixed	mixed	4	5000	63	5.07	6.17	99.1	22.7	53.1	-0.033	0.000	0.98	1.01
strong	weak	4	5000	37	4.70	4.36	99.7	0.2	72.6	-0.061	-0.000	1.05	1.02
strong	strong	4	5000	119	5.77	10.00	97.5	0.0	29.7	-0.109	-0.001	1.08	1.01

A.4. Monte Carlo iterative processes to determine parameter values for the outcome and treatment-selection logistic models

In this section, we describe an iterative process using Monte Carlo integration (Austin, A Data-Generation Process for Data with Specified Risk Differences or Numbers Needed to Treat, 2010) to determine the logistic model parameter values $\alpha_{0,treat}$, $\alpha_{0,outcome}$, and β that will result in the desired proportion of subjects exposed to the experimental treatment, the true probability of the outcome if all experimental subjects were under the control treatment, and the true risk difference for the experimental treated population (i.e. ATT), respectively. We will first describe the process of determining the value of $\alpha_{0,treat}$. First, we define an interval [$\alpha_{0,treat}^{lower}$, $\alpha_{0,treat}^{upper}$] that contains the value of $\alpha_{0,treat}$ that results in the desired proportion of subjects exposed to the experimental treatment. The choice of $\alpha_{0,treat}^{lower}$ and $\alpha_{0,treat}^{upper}$ can be chosen by trial and error. The iterative process is begun by setting $\alpha_{0,treat}^{(1)}$ to be equal to the mid-point of the interval [$\alpha_{0,treat}^{lower}$, $\alpha_{0,treat}^{upper}$], where $\alpha_{0,treat}^{(1)}$ is the value of $\alpha_{0,treat}$ in the first iteration. Then, we simulate a data set under a specified scenario (e.g. independent standard normal covariates) and we compute the probability of being exposed to the experimental treatment, for each subject, based on a specified treatment-selection logistic model described in section 3.1, with $\alpha_{0,treat} = \alpha_{0,treat}^{(1)}$. We then compute the mean probability of being exposed to the experimental treatment across the simulated population. This process is then repeated 1,000 times, and the mean probability of being exposed to the experimental treatment is determined across the 1,000 simulated datasets. If the mean probability of being exposed to the experimental treatment across the 1,000 simulated datasets is greater than the desired proportion of subjects exposed to the experimental treatment, then $\alpha_{0,treat}^{(2)}$ is chosen to be the mid-point of the interval $[\alpha_{0,treat}^{lower}, \alpha_{0,treat}^{(1)}]$, otherwise $\alpha_{0,treat}^{(2)}$ is chosen to be the mid-point of the interval $[\alpha_{0,treat}^{(1)}, \alpha_{0,treat}^{upper}]$, where $\alpha_{0,treat}^{(2)}$ is the value of $\alpha_{0,treat}$ in the second iteration. This process is repeated until the mean probability of being exposed to the experimental treatment across a 1,000 simulated datasets is arbitrarily close (i.e. converges) to the desired proportion of subjects exposed to the experimental treatment.

We will now describe the process of determining the intercept in the outcome-selection logistic model ($\alpha_{0,outcome}$). First, we define an interval $[\alpha_{0,outcome}^{lower}, \alpha_{0,outcome}^{upper}]$ that contains the value of $\alpha_{0,outcome}$ that results in the desired probability of the outcome if all experimental subjects were under the control treatment. The choice of $\alpha_{0outcome}^{lower}$ and $\alpha_{0,outcome}^{upper}$ can be chosen by trial and error. The iterative process is begun by setting $\alpha_{0,outcome}^{(1)}$ to be equal to the mid-point of the interval $[\alpha_{0,outcome}^{lower}, \alpha_{0,outcome}^{upper}]$, where $\alpha_{outcome}^{(1)}$ is the value of $\alpha_{0,outcome}$ in the first iteration. Then, for each subject, we generate values for the covariates under a specified covariate scenario and we generate a binary treatment status indicator (i.e. T_i=1 for experimental and 0 for control) based on a specified treatment-selection logistic model described in section 3.1. Next, for each subject, we compute the probability of the outcome under the control based on a specified outcome-selection logistic model described in section 3.1, with $\alpha_{0,outcome} = \alpha_{0,outcome}^{(1)}$. Then, we compute the mean of these probabilities across those exposed to the experimental treatment (T_i=1). This process is then repeated 1,000 times, and the mean probability of the outcome under the control for those exposed to the experimental treatment is determined across the 1,000 simulated datasets. If the mean probability of the outcome under the control for experimental treated subjects across the 1,000 simulated datasets is greater than the desired event probability , then $\alpha_{0,outcome}^{(2)}$ is chosen to be the mid-point of the interval [$\alpha_{0,outcome}^{lower}$, $\alpha_{0,outcome}^{(1)}$], otherwise $\alpha_{0,outcome}^{(2)}$ is chosen to be the mid-point of the interval [$\alpha_{0,outcome}^{(1)}$, $\alpha_{0,outcome}^{upper}$], where $\alpha_{0,outcome}^{(2)}$ is the value of $\alpha_{0,outocme}$ in the second iteration. This process is repeated until the mean probability of the outcome under the control for experimental treated subjects is arbitrarily close (i.e. converges) to the desired event probability.

Lastly, we will now describe the process of determining the coefficient for the treatment variable in the outcome-selection logistic model (β). First, we define an interval [β^{lower}, β^{upper}] that contains the value of β that results in the desired risk difference for the experimental treated population. The choice of $\beta^{lower}$and β^{upper} can be chosen by trial and error. The iterative process is begun by setting $\beta^{(1)}$ to be equal to the mid-point of the interval [β^{lower}, β^{upper}] where $\beta^{(1)}$ is the value of β in the first iteration. Then, for each subject, we generate values for the covariates under a specified covariate scenario and we generate a binary treatment status indicator (i.e. T_i=1 for experimental and 0 for control) based on a specified treatment-selection logistic model described in section 3.1. Next, for each subject, we compute the probability of the outcome under the control treatment and the probability of the outcome under the experimental treatment based on a specified outcome-selection logistic model described in section 3.1, with $= \beta^{(1)}$. Then, for each subject, we compute the risk difference (i.e. difference between the probability of the outcome under the control and the probability of the outcome under the experimental). Next, we compute the mean risk difference across the experimental subjects (T_i=1). This process is then repeated 1,000 times, and the mean risk difference for those exposed to the experimental treatment is determined across the 1,000 simulated datasets. If the mean risk difference across the 1,000 simulated datasets is greater than the desired risk difference, then $\beta^{(2)}$ is chosen to be the mid-point of the interval [$\beta^{lower}, \beta^{(1)}$], otherwise $\beta^{(2)}$ is chosen to be the mid-point of the interval [$\beta^{(1)}, \beta^{upper}$,], where $\beta^{(2)}$ is the value of β in the second iteration. This process is repeated until the mean risk difference for the experimental treated subjects across 1000 simulated datasets is arbitrarily close (i.e. converges) to the desired risk difference.

BIBLIOGRAPHY

Austin, P. C. (2008). A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in Medicine*, 27: 2037-2049.

Austin, P. C. (2008). Assessing balance in measured baseline covariates when using many-to-one matching on the propensity-score. *Pharmacoepidemiology and drug safety, 17*, 1218-1225.

Austin, P. C. (2009). Type I error rates, coverage of confidence intervals, and variance estimation in propensity-score matched analyses. *The International Journal of Biostatistics, 5*(13).

Austin, P. C. (2010). A Data-Generation Process for Data with Specified Risk Differences or Numbers Needed to Treat. *Communications in Statistics-Simulation and Computation, 39*, 563-577.

Austin, P. C. (2010). Statistical Criteria for Selecting the Optimal Number of Untreated Subjects Matched to Each Treated Subject When Using Many-to-One Matching on the Propensity Score. *American Journal of Epidemiology, 172*, 1092-1097.

Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research, 46*, 399-424.

Austin, P. C. (2011). Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples. *Statistics in Medicine, 30*, 1292-1301.

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics, 10*, 150-161.

Boening, A., Friedrich, C., Hedderich, J., Schoettler, J., Fraund, S., & Cremer, J. T. (2003). Early and Medium-Term Results After On-Pump and Off-Pump Coronary Artery Surgery: A Propensity Score Analysis. *The Annals of Thoracic Surgery, 76*(6), 2000–2006.

Coca-Perraillon, M. (2007). *Local and Global Optimal Propensity Score Matching.* SAS Global Forum 2007 - Statistics and Data Analysis.

Dann, R. S., & Koch, G. G. (2008). Methods for one-sided testing of the difference between proportions and sample size considerations related to non-inferiority clinical trials. *Pharmaceutical Statistics, 7*, 130-141.

Deb, S., Austin, P. C., Tu, J. V., Ko, D. T., Mazer, C., Kiss, A., & Fremes, S. E. (2016). A review of propensity-score methods and their use in cardiovascular research. *Canadian Journal of Cardiology, 32*, 259-265.

Durkalski, V. L., Palesch, Y. Y., Lipsitz, S. R., & Rust, P. F. (2003). Analysis of clustered matched-pair data for a non-inferiority study design. *Statistics in Medicine, 22*, 279-290.

Eldridge, S. M., Ashby, D., & Kerry, S. (2006). Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *International Journal of Epidemiology , 35*, 1292-1300.

Eliasziw, M., & Donner, A. (1991). Application of the McNemar test to non-independent matched-pair data. *Statistics in Medicine, 10*, 1981-1991.

Farrington, C. P., & Manning, G. (1990). Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine*, 9, 1447-1454.

Food and Drug Administration. (2013). *Guidance for Industry and FDA Staff Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets.* Retrieved 1 30, 2017, from http://www.fda.gov/downloads/drugs/guidances/ucm243537.pdf

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*, 199-236.

Imbens, G. W. (2004). Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. *The Review of Economics and Statistics, 86*(1), 4-29.

Jiang, Q., & Xia, H. (2015). *Quantitative evaluation of safety in drug development: design, analysis, and reporting.* Boca Raton: CRC Press Taylor and Francis Group.

Kereiakes, D. J., Yeh, R. W., Massaro, J. M., Driscoll-Shempp, P., & et al. (2015). Stent Thrombosis in Drug-Eluting or Bare-Metal Stents in Patients Receiving Dual

Antiplatelet Therapy. *Journal of the American College of Cardiology: Cardiovascular Interventions, 8*(12), 1552-62.

Ko, D. T., Chiu, M., Austin, P. C., Bowen, J., Cohen, E. A., & Tu, J. V. (2008). Safety and effectiveness of drug-eluting stents among diabetic patients: A propensity analysis. *American Heart Journal, 156*, 125-34.

Ko, D. T., Chiu, M., Guo, H., Austin, P. C., Goeree, R., Cohen, E., . . . Tu, J. V. (2009). Safety and Effectiveness of Drug-Eluting and Bare-Metal Stents for Patients with off- and on-label indications. *Journal of the American College of Cardiology, 53*(19), 1773-82.

Lee, S. S. (2008). *Analysis of correlated binary data in non-inferiority trials.* Boston: Boston University, ProQuest Dissertations.

Levenson, M. S., & Yue, L. Q. (2013). Regulatory issues of propensity score methodology application to drug and device safety studies. *Journal of Biopharmaceutical Statistics, 23*, 110-121.

Mascha, E. J., & Sessler, D. I. (2011). Equivalence and Noninferiority Testing in Regression Models and Repeated-Measures Designs. *Anesthesia and Analgesia*, Vol. 112(3), 678-687.

Mauri, L., Kereiakes, D. J., Yeh, R. W., & al., e. (2014). Twelve or 30 months of dual anti-platelet therapy after drug-eluting stents. *New England Journal of Medicine, 371*, 2155-66.

Ming, K., & Rosenbaum, P. R. (2000). Substantial gains in bias reduction from matching with a variable number of controls. *Biometrics, 56(1)*, 56, 118-124.

Nam, J.-m. (1997). Establishing equivalence of two treatments and sample size requirements in matched-pair data. *Biometrics, 53*, 1422-1230.

Nam, J.-m., & Kwon, D. (2009). Non-inferiority tests for clustered matched pair data. *Statistics in Medicine, 28*, 1668-1679.

Obuchowski, N. A. (1998). On the comparison of correlated proportions for clustered data. *Statistics in Medicine, 17*, 1495-1507.

Pyo, J. H., Lee, H., Min, B.-H., Lee, J. H., & et al. (2016). Long-Term Outcome of Endoscopic Resection vs. Surgery for Early Gastric Cancer: A Non-inferiority Matched Cohort Study. *American Journal of Gastroenterology, 111*, 240-249.

Rassen, J. A., Shelat, A. A., Myers, J., Glynn, R. J., Rothman, K. J., & Schneeweiss, S. (2012). One-to-many propensity score matching in cohort studies. *Pharmacoepidemiology and Drug Safety , 21*, 69-80.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*(1), 33-38.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology, 66*(5), 688-701.

SAS. (2012, July 16). *Usage Note 46997: Estimating the risk (proportion) difference for matched pairs data with binary response*. (SAS) Retrieved 2016, from SAS-Support: Samples & SAS Notes: http://support.sas.com/kb/46/997.html

Schafer, J. L., & Kang, J. D. (2008). Average causal effects from nonrandomized studies: a practical guide and simulated example. *Psychological Methods, 13*, 279-313.

Spiegelman, D., & Hertzmark, E. (2005). Easy SAS Calculations for Risk or Prevalence Ratios and Differences. *American Journal of Epidemiology*, Vol. 162(3), 199-200.

Stuart, E. A. (2008). Developing practical recommendations for the use of propensity scores: Discussion of 'A critical appraisal of propensity score matching in the medical literature between 1996 and 2003' by Peter Austin, Statistics in Medicine. *Statistics in Medicine, 27*, 2062-2065.

Stuart, E. A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science, 25*, 1-21.

Wu, S., Crespi, C. M., & Wong, W. K. (2012). Comparison of Methods for Estimating the Intraclass Correlation Coefficient for Binary Responses in Cancer Prevention Cluster Randomized Trials. *Contemporary Clinical Trials, 33*, 869-880.

Yang, Z., Sun, X., & Hardin, J. W. (2012). Testing non-inferiority for clustered matched-pair binary data in diagnostic medicine. *Computational Statistics and Data Analysis, 56*, 1301-1320.

CURRICULUM VITAE

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]